Investigation on the Efficient Frontier Based on CVaR under Copula Dependence Structure with Applications to South African JSE Stocks

W.B.Damaseb

A thesis submitted in partial fulfillment

of the requirements for the degree of

M.Sc. Financial Mathematics

University of Cape Town

2005

Internal Supervisor: Dr. P. Ouwehand
Department of Mathematics and Applied Mathematics

External Supervisor: Dr. A. Demchuk

November 16, 2005

Abstract

University of Cape Town Abstract

Investigation on the Efficient Frontier Based on CVaR under Copula Dependence Structure with Applications to South African JSE Stocks

By: W. B. Damaseb

Internal Supervisor: Dr. P. Ouwehand
Department of Mathematics and Applied Mathematics

External Supervisor: Dr. A. Demchuk

We study the feasibility of using a coherent monetary risk measure, Conditional Value at Risk (CVaR) also known as Expected Shortfall (ES), to optimise a portfolio of South African stocks. Value at Risk (VaR) is not a sub-additive risk measure and therefore does not possess one of the four properties that all coherent risk measures must satisfy. Using copula to describe the dependence structure between the instruments in our portfolio, we implement and backtest a CVaR optimization algorithm and compare the backtested results to those obtained using parametric and non-parametric/Monte Carlo VaR. Finally we optimise the portfolio of stocks and generate an efficient frontier specifying CVaR as the risk measure instead of the portfolio variance traditionally used in Markowitz and CAPM models. The errors and tracking errors for the returns of the CVaR, Markowitz and CAPM frontier portfolios to the actual portfolio returns are then compared.

Declaration

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.

2. I have used the IEEE scientific convention for citation and referencing. Each significant contribution to, and quotation in this thesis from the work(s) of other people has been attributed, and has been cited and referenced.

3. This thesis is my own work.

4. I have not allowed, and will not allow anyone to copy my work with the intention of passing it off as his or her own work.

Signature ....

Acknowledgments

This study could not have been possible without the help/input of a few individuals.

- Dr. Peter Ouwehand for proposing the topic.
- Dr. Andriy Demchuck for his guidance and support.
- Mr John Kyeyune for his invaluable brainstorming sessions.
- All my family and friends.

Terms of Reference

The following study was commissioned by Dr. P.Ouwehand from the Department of Mathematics at the University of Cape Town. And the instructions were as follows:

- Familiarise yourself with basic portfolio theory: Markowitz, CAPM;
- Find out about the weaknesses of these models;
- Familiarise yourself with Value-at-Risk and the literature on coherent and convex measures;
- Develop, Investigate and Implement portfolio optimization techniques that use say, VaR or expected shortfall risk;
- Assess your models performance (using data or simulation);
- Draw conclusions on the findings and;
- Submit the study.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Capital preservation is not a new concern for investors and regulators. No one can assure an investor that their initial investment will generate a profit every day, for the period they are invested, and this may cause a few investors to be sceptical about handing over their hard-earned money if their worst expected loses cannot be quantified.

Regulators on the other hand need a method to determine capital requirements for bankers, traders and regular market participants to regulate the market risk assumed by these participants. Value at Risk (VaR) has been widely used as a risk measure statistic by regulators to determine capital requirements.

Coherent Measures of Risk [1] introduce four desirable properties for a sensible risk measure, naming measures that satisfy these properties coherent. As discussed by [1], [2] and [3] VaR does not fall into this category of coherent risk measures due to the fact that it does not posses the property of sub-additivity. Consequently, VaR does not recognise the diversification effects of a portfolio and as an alternative to VaR a coherent risk measure called Expected Shortfall (ES) or Conditional Value at Risk (CVaR) is introduced.

Gaivoronski and Pflug [4] compare portfolio optimization methods for VaR and CVaR. They find that generating efficient frontiers with CVaR as the risk measure are easier to compute than generating frontiers using VaR as the risk measure, due to the non-convex nature of VaR generating many local minima.

Portfolio optimization algorithms for CVaR are also discussed in [25], [5], [23] and [21].

Traditional methods for optimizing the risk-return profile of a portfolio in modern portfolio theory use variance as the risk measure, examples of such methods are the CAPM and Markowitz [9] model.

Tail events are very important in measuring VaR and CVaR and there has been criticism about using conditional multivariate normal distributions to describe the dependence structure between instruments in a portfolio. A full investigation of this topic is beyond the scope of this work, but [17], [14] and [13] find that even though there is still some research to be conducted, copulas describe the dependence structure between instruments in a portfolio better than multivariate normal distributions.

The main aim of this thesis is to investigate the feasibility of using monetary risk measures for portfolio optimization. Examples of monetary risk measures are Value at Risk, coherent risk measures and convex risk measures. These risk measures approach risk from a capital adequacy and regulatory point of view. Instead of asking, "How risky is this investment?" the question is, "How much capital should be set aside to hedge down-side risk?"

The scope of the study is limited to the investigation of Conditional Value at Risk as the monetary risk measure we seek to optimise for our portfolio. This monetary risk measure is back-tested to determine its accuracy and compared to the accuracy of a VaR backtest.

We also utilise the CVaR algorithm to obtain optimal portfolios for given risk levels, where CVaR acts as the measure of risk. The actual out of period portfolio returns from the CVaR efficient frontier are compared to those obtained using CAPM and Markowitz efficient frontiers.

We are interested in setting up an optimization model that is easy to modify, programmed in a package that is commercially available to most computer users and is not computationally intensive.

The data for the historical/backtesting analysis of the model is selected from the *Troskie*[1] database. The *Troskie* database has monthly and weekly returns for South African stocks, to have as many data points as possible we select the weekly data to perform our analysis.

The thesis outline is as follows:

We start with a review of modern portfolio theory by discussing the Markowitz and CAPM models, the assumptions and underlying theory for each model are reviewed and the disadvantages of these models are also mentioned.

The next section is an introduction to Value at Risk (VaR) and Conditional Value at Risk (CVaR) also known as mean shortfall or tail VaR. We start by discussing parametric and Non-parametric VaR, this is followed by giving justification for using a coherent risk measure and we give an example as to why VaR is not a sub-additive risk measure, meaning that the VaR of a portfolio of two instruments can be higher than the sum of the individual VaR of each instrument. The section ends by introducing the underlying theory and properties of CVaR including the CVaR model that we will use to optimise our portfolio.

Dependence concepts in financial risk management are discussed and we introduce the use of copulas, specifically the t-copula as our choice of dependence structure for the instruments in our portfolio. Following this section is a discussion on how to model events in practice concentrating on parameter estimation and generating random numbers.

The above-mentioned theory is put into practice with a case study on the South African market; six stocks[2] are selected from the South African stock mar-

[1] University of Cape Town Statistics Department

[2] We could only create an exhaustive database for these 6 stocks, so that there are no gaps in the time series of historical prices

ket as our portfolio. This portfolio is then used to backtest the CVaR model for a 95% and 99% confidence level, next we generate efficient frontiers for our CVaR model and the Markowitz and CAPM models and calculate the tracking errors of projected returns to actual portfolio returns had we invested in these efficient frontier portfolios.

The results obtained from the case study show that the backtested CVaR model performs better for the 95% and 99% confidence levels than the parametric and Non-parametric VaR models when backtested. When compared to the CAPM and Markowitz models, the CVaR model efficient frontier expected returns have a higher tracking error[3] to the actual portfolio returns, when invested in the frontier portfolios, but significantly less negative errors[4]. The study then ends with recommendations for future work on the CVaR model.

[3]Standard deviation of the difference between the expected returns from optimisation and actual portfolio returns.

[4]Difference between expected returns from optimisation and actual portfolio returns with optimised allocations.

Chapter 2

Modern Portfolio Theory

2.1 Markowitz Portfolio Optimization

This section follows the discussion of [9]. Harry Markowitz suggested that for any given level of risk (volatility), the rational investor would select the maximum return and for a given level of return, the rational investor would select the minimum risk. The main assumptions underlying the model can be found in almost any text on modern portfolio theory [18]:

- The final outcome of an investment is summarised by its return, and investors use a probability distribution of the rates of return of an investment;
- The Investors risk estimates for an investment are proportional to the variance of return for the investment;
- Investors base their decisions on just two parameters of the probability distribution function, being the expected return and variance of the return;
- All investors are risk averse, in other words (as mentioned before) for any given level of risk (volatility), the rational investor would select the maximum return and for a given level of return, the rational investor would select the minimum risk.

For a portfolio of n securities the expected return of the portfolio is the weighted average of the expected return of each security. If we let the vector of n security returns be

$$\mathbf{R} = \begin{pmatrix} R_1 \\ . \\ . \\ . \\ R_n \end{pmatrix}$$

so that $E(R) = \mu$.

$$\mu = \begin{pmatrix} \mu_1 \\ . \\ . \\ . \\ \mu_n \end{pmatrix}$$

With covariance matrix $\Sigma = E(R-\mu)(R-\mu)'$. If we then assume multivariate normality $\mathbf{R} \sim N(\mu, \Sigma)$ and let w_i be the weight of security i such that

$$\sum_{i=1}^{n} w_i = 1 \tag{2.1}$$

and

$$\mathbf{W} = \begin{pmatrix} w_1 \\ . \\ . \\ . \\ w_n \end{pmatrix}$$

Then the expected return for the portfolio is

$$\mu_P = E(R_P) = \mathbf{W}'E(\mathbf{R}) = \sum_{i=1}^{n} w_i \mu_i \tag{2.2}$$

where R_P is the return of the portfolio, w_i is the proportion of security i in the portfolio and μ_i is the expected return of security i. The variance of the portfolio is calculated using the variance of return of each security and the covariance of returns between each pair of securities.

$$\sigma_P^2 = Var(R_P) = \mathbf{W}'\Sigma\mathbf{W} = \sum_{i,j=1}^{n} w_i w_j \sigma_{ij} \tag{2.3}$$

σ_{ij} is the covariance between the returns of securities i and j and depends on the correlation between the returns of the two securities $\sigma_{ij} = \rho_{ij}\sigma_i\sigma j$, where

ρ_{ij} is the correlation coefficient between the returns of securities i and j.

Diversification of a portfolio reduces the risk/variance of the portfolio if the securities are not perfectly correlated. To illustrate this point let us consider a portfolio with two securities such that $\sigma_1=\sigma_2=\sigma$, the variance of the portfolio is calculated as follows.

$$\begin{aligned} Var(R_P) &= w_1^2\sigma^2 + (1-w_1)^2\sigma^2 + 2w_1(1-w_1)\rho_{12}\sigma^2 \\ &= \sigma^2[w_1^2 + (1-w_1)^2 + 2w_1(1-w_1)\rho_{12}] \end{aligned} \tag{2.4}$$

If $\rho_{12}=1$ then $w_1^2 + (1-w_1)^2 + 2w_1(1-w_1) = 1$ and equation (2.4) becomes:

$$Var(R_P) = \sigma^2$$

If $\rho_{12} < 1$ then

$$Var(R_P) \leq \sigma^2$$

Low correlations between the securities therefore lead to the reduction of the variance of the portfolio.

The Markowitz optimization does not produce one optimal portfolio, it provides a group of optimal portfolios with security weights determined by different risk/return characteristics. These optimal portfolios make up what is called the efficient frontier and the investor chooses from this frontier according to their risk/return preferences. The Markowitz portfolio algorithm is as follows.

$$\begin{aligned} \max_{w_i} E(P) &= \mathbf{W}'\mu \\ &= \sum_{i=1}^{n} w_i\mu_i \\ \min_{w_i} \sigma_P^2 &= \mathbf{W}'\Sigma\mathbf{W} \\ &= \sum_{i,j=1}^{n} w_i w_j \sigma ij \ \ subject\ to \\ \sum_{i=1}^{n} w_i &= 1, 0 \leq w_i \leq 1, i = 1, ..., n \end{aligned} \tag{2.5}$$

This is a quadratic programming (QP)problem, in order to solve it we can fix the return μ_P and then minimise the variance σ_P^2 or fix the variance and maximise the return. If we choose to fix the desried percentage return, $\mu_p = A$, we need to solve the following QP problem.

$$\begin{aligned}
\min_{w_i} \sigma_P^2 &= \mathbf{W}'\Sigma\mathbf{W} \\
&= \sum_{i,j=1}^{n} w_i w_j \sigma ij, \quad \textit{subject to} \\
\mu_P &= \mathbf{W}'\mu = \sum_{i=1}^{n} w_i \mu_i = A, \quad \textit{subject to} \\
\sum_{i=1}^{n} w_i &= 1, 0 \leq w_i \leq 1, i = 1, ..., n
\end{aligned}$$

Varying A will yield the efficient frontier for the specific portfolio.

A few of the main problems arising from using Markowitz's approach of optimization are:

- The assumption of a multivariate normal distribution for the instruments of the portfolio is not correct, [7], [17], [22];
- The portfolio optimization is based entirely on historical results and is therefore backward looking;
- Due to the backward looking nature of the optimization method stress testing is not possible.

2.1.1 Sharpe's corner portfolio

Sharpe introduced the concept of corner portfolios to reduce Markowitz's QP problem to a Linear programming (LP) problem. Consider the efficient frontier plot of return μ_P versus variance σ_P^2. If we draw a line through the frontier $\mu_P = A + B\sigma_P^2$, then.

$$\begin{aligned}
\mu_P &= A + B\sigma_P^2 \\
A &= \mu_P - B\sigma_P^2 \\
\frac{A}{B} &= \frac{1}{B}\mu_P - \sigma_P^2 \\
Z &= \theta\mu_P - \sigma_P^2 \qquad (2.6)
\end{aligned}$$

If we let the slope $B = \frac{1}{\theta}$ then the line parallel to the $\sigma_P^2 = X$ axis is where the slope $B = 0$ and $\theta = \infty$, if we then maximise Z in equation (2.6) keeping θ fixed (maximise A keeping B fixed) this will give us a point on the efficient frontier, varying θ from $(0, \infty)$ produces the efficient frontier, and our optimization problem reduces to.

$$\begin{aligned}
Max(Z) &= \theta\mu_P - \sigma_P^2 \\
&= \theta\mathbf{W}'\mu - \mathbf{W}'\Sigma\mathbf{W} \textit{ subject to} \qquad (2.7) \\
\sum_{i=1}^{n} w_i &= 1
\end{aligned}$$

w_i is not restricted to positive values, which means short selling (leverage) is allowed. using Legrange multipliers to include the constraints in the function Z to be maximised the problem becomes.

$$\begin{aligned}
Max(Z') &= \theta\mathbf{W}'\mu - \mathbf{W}'\Sigma\mathbf{W} + \lambda(1 - \sum_{i=1}^{n} w_i) \\
&= \theta\sum_{i=1}^{n} w_i\mu_i - \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{ij} + \lambda(1 - \sum_{i=1}^{n} w_i) \qquad (2.8)
\end{aligned}$$

Taking partial derivatives of Z' with respect to $w_i, i = 1, ..., n$ and λ and equating them to zero we get the following simultaneous equations.

$$
\begin{aligned}
\frac{\partial Z'}{\partial w_1} &= \theta\mu_1 - 2w_1\sigma_{11} - 2w_2\sigma_{12} - \ldots - 2w_n\sigma_{1n} - \lambda = 0 \\
\frac{\partial Z'}{\partial w_2} &= \theta\mu_2 - 2w_2\sigma_{22} - 2w_1\sigma_{21} - \ldots - 2w_n\sigma_{2n} - \lambda = 0 \\
\vdots &= \vdots \\
\frac{\partial Z'}{\partial w_n} &= \theta\mu_n - 2w_n\sigma_{nn} - 2w_1\sigma_{n1} - \ldots - 2w_{n-1}\sigma_{n,n-1} - \lambda = 0 \\
\frac{\partial Z'}{\partial \lambda} &= 1 - w_1 - w_2 - \ldots - w_n = 0
\end{aligned}
$$

Converted to matrix notation this yields.

$$
\begin{pmatrix}
2\sigma_{11} & 2\sigma_{12} & \ldots & 2\sigma_{1n} & 1 \\
2\sigma_{21} & 2\sigma_{22} & \ldots & 2\sigma_{2n} & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
2\sigma_{n1} & 2\sigma_{n2} & \ldots & 2\sigma_{nn} & 1 \\
1 & 1 & \ldots & 1 & 0
\end{pmatrix}
\cdot
\begin{pmatrix}
w_1 \\ w_2 \\ . \\ . \\ . \\ w_n \\ \lambda
\end{pmatrix}
=
\begin{pmatrix}
\theta\mu_1 \\ \theta\mu_2 \\ . \\ . \\ . \\ \theta\mu_n \\ 1
\end{pmatrix}
$$

$$
\begin{aligned}
GX &= B \\
X &= G^{-1}B \textit{ so that} \\
w_i &= c_i + e_i(\theta), i = 1, \ldots, n \\
\lambda &= c_\lambda + e_\lambda(\theta)
\end{aligned}
$$

Where c_i and c_λ are the constant components of w_i and λ, $e_i(\theta)$ and $e_\lambda(\theta)$ are a function of the varying θ for w_i and λ respectively, so for different values of θ the expected return and variance of the portfolio is.

$$
\begin{aligned}
\mu_P &= \mathbf{W}'\mu \\
&= \sum_{i=1}^{n} w_i\mu_i \\
&= \sum_{i=1}^{n} (c_i + e_i(\theta))\mu_i \\
&= \sum_{i=1}^{n} c_i\mu_i + \theta\sum_{i=1}^{n} e_i\mu_i
\end{aligned}
\tag{2.9}
$$

$$
\begin{aligned}
Var(P) &= \sigma_P^n = \mathbf{W}'\Sigma\mathbf{W} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{ij} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} (c_i + e_i(\theta))(c_j + e_j(\theta))\sigma_{ij} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \sigma_{ij} + \theta \sum_{i=1}^{n}\sum_{j=1}^{n} (c_i e_j + c_j e_i)\sigma_{ij} + \\
&\quad \theta^2 \sum_{i=1}^{n}\sum_{j=1}^{n} e_i e_j \sigma_{ij} \qquad (2.10)
\end{aligned}
$$

Given the covariance between the returns of the securities σ_{ij} and the expected return μ_i of each security. We can vary θ from 0 to ∞ and calculate the set of weights $\mathbf{W}$ which is optimum for each value of θ for each security.

If we wish to eliminate short selling, as soon as a securities weight becomes 0 we eliminate this security from the portfolio and re-evaluate the weights at the corresponding θ again, only varying θ if no other security has 0 weighting.

Although equation (2.7) still looks like a QP, Sharpe varies θ from ∞ to 0, and since returns and variances are small values, the term $\theta\mu_P$ dominates (2.7) and we can ignore σ_P^2 and the Sharpe formulation becomes.

$$
\begin{aligned}
Max(Z) &= \theta\mu_P \\
&= \theta\mathbf{W}'\mu \\
&= \theta\sum_{i=1}^{n} w_i\mu_i \quad \textit{subject to} \\
\sum_{i=1}^{n} w_i &= 1
\end{aligned}
$$

This is a LP problem which can be easily solved using a PC. So once a starting value for θ is chosen this algorithm is used to generate the efficient frontier by generating the corner portfolios, which is a portfolio where a security either enters or leaves a portfolio as θ is varied. Equations (2.9) and (2.10) can then be used to calculate the variance σ_P and expected return $E(P)$ for the portfolio.

Example: Let $E(R_1) = \mu_1 = 0.10$, $E(R_2) = \mu_2 = 0.15$, $\sigma_1 = 0.3$, $\sigma 2 = 0.4$ and $\rho_{12} = 0$ so that $\sigma_{11} = 0.09$, $\sigma_{22} = 0.16$ and $\sigma_{12} = 0$. Our system of equations then becomes $AX = B$.

$$\begin{pmatrix} 0.18 & 0 & 1 \\ 0 & 0.32 & 1 \\ 1 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ \lambda \end{pmatrix} = \begin{pmatrix} 0.10\theta \\ 0.15\theta \\ 1 \end{pmatrix}$$

So that $X = A^{-1}B$

$$\begin{pmatrix} w_1 \\ w_2 \\ \lambda \end{pmatrix} = \begin{pmatrix} 2.00 & -2.00 & 0.64 \\ -2.00 & 2.00 & 0.36 \\ 0.64 & 0.36 & -0.12 \end{pmatrix} \cdot \begin{pmatrix} 0.10\theta \\ 0.15\theta \\ 1 \end{pmatrix}$$

With solutions.

$$\begin{aligned} w_1 &= 2.00(0.1\theta) - 2.00(0.15\theta) + 0.64 = 0.64 - 0.10\theta \\ w_2 &= -2.00(0.1\theta) + 2.00(0.15\theta) + 0.36 = 0.36 + 0.10\theta \\ \lambda &= 0.64(0.1\theta) + 0.36(0.15\theta) - 0.12 = -0.12 + 0.12\theta \end{aligned} \tag{2.11}$$

Varying θ from 0 to 10 we get.

$$\begin{aligned} \theta &= 0 \begin{bmatrix} w_1 = 0.64 \\ w_2 = 0.36 \end{bmatrix}, E(P) = \mu_p = 11.80\% \\ \theta &= 0.5 \begin{bmatrix} w_1 = 0.59 \\ w_2 = 0.41 \end{bmatrix}, E(P) = \mu_p = 12.05\% \\ \theta &= 1 \begin{bmatrix} w_1 = 0.54 \\ w_2 = 0.46 \end{bmatrix}, E(P) = \mu_p = 12.30\% \\ \theta &= 2 \begin{bmatrix} w_1 = 0.44 \\ w_2 = 0.56 \end{bmatrix}, E(P) = \mu_p = 12.80\% \\ \theta &= 5 \begin{bmatrix} w_1 = 0.14 \\ w_2 = 0.86 \end{bmatrix}, E(P) = \mu_p = 14.30\% \\ \theta &= 10 \begin{bmatrix} w_1 = -0.36 \\ w_2 = 1.36 \end{bmatrix}, E(P) = \mu_p = 16.80\% \end{aligned}$$

2.2 Capital asset Pricing Model (CAPM)

The CAPM model section follows the discussion of [18]. The simplifying assumptions underlying the CAPM model are [12], [18]:

1. All investors use the Markowitz portfolio selection model;
2. There are many investors, each with a wealth that is small compared to the total wealth of all the investors combined and investors act as though security prices are unaffected by their own trades, this leads to the perfect competition assumption of microeconomics;
3. All investors plan for the same holding period for their portfolios, which is a suboptimal, since we ignore all information after the single period horizon;
4. No transaction costs, interest rate changes, Inflation and taxes;
5. Investments are limited to publicly traded financial assets such as: stocks, bonds and unlimited risk-free borrowing/lending. This assumption means that private enterprises and government funded assets such as international airports may not be invested in;
6. All investors share the same economic view of the world, so that all investors use the same expected returns and covariance matrix of security returns to generate the efficient frontier and optimal risky portfolio.

The Market Model

Let a portfolio M (market portfolio) represent the total economy, such that it is the weighted average of all the quoted securities, investing in this theoretical portfolio should guarantee an expected return of the market, R_M. Portfolio M is important because it is a perfectly diversified portfolio.

Sharpe's single index model assumes that a securities price movements can be related to a market index by a parameter β. In order to calculate β we need to find a benchmark representing the economy, such as portfolio M or some Index. Using this benchmark and linear regression it is possible to measure the change in security returns as relates to changes in general economic conditions, known as the market model. β therefore indicates the average sensitivity of returns of a security to the market return.

Using least squares estimation, the market model is given by.

$$R_i = \alpha_i + \beta_i R_M + U_i \tag{2.12}$$

The α_i indicates that companies do not only earn returns associated with general movements in the economy, but also specific returns associated with the companies individual activities. U_i is a random error term, $U_i \sim N(0,1)$ and helps determine the specific return alongside the α_i term, the error terms are independent of each other $(Cov(U_i, U_j) = 0, \forall i, j, i \neq j)$and R_M $(Cov(U_i, R_M) = 0, \forall i)$. The market model provides the conceptual foundation for the CAPM model.

The Capital Market Line (CML) and Risk-Free Assets

Let us consider a two-asset risky portfolio, with portfolio weights w_1, w_2, returns μ_1, μ_2 and variance σ_1^2, σ_2^2 for the assets 1 and 2 respectively and correlation coefficient ρ_{12}. The expected rate of return and variance of the risky portfolio is given by.

$$\begin{aligned} E(R_r) &= w_1\mu_1 + w_2\mu_2 \qquad (2.13) \\ w_2 &= 1 - w_1 \end{aligned}$$

$$Var(r) = w_1^2\sigma_1^2 + (1-w_1)^2\sigma_2^2 + 2w_1(1-w_1)\rho_{12}\sigma_1\sigma_2 \tag{2.14}$$

If we add a risk-free asset R_F to our portfolio, and give a weighting of w to our risky portfolio then the expected rate of return and variance of this new portfolio is.

$$E(R_p) = wE(R_r) + (1-w)R_F \tag{2.15}$$

$$Var(p) = w^2\sigma^2 \tag{2.16}$$

Where σ^2 is the variance of the risky portfolio, $Var(r)$, this is because the variance of the risk-free asset is zero and the correlation coefficient between the risk-free asset and risky portfolio is not defined. The standard deviation of the combined portfolio is therefore.

$$Sdev(p) = w\sigma \tag{2.17}$$

If we solve for w in equation (2.15) and substitute it in equation (2.17) we get.

$$E(R_p) = R_F + \frac{E(R_r) - R_F}{\sigma} Sdev(p) \tag{2.18}$$

Equation (2.18) indicates a linear relationship between $E(R_p)$ and $Sdev(p)$, with an intercept at R_F and a slope of $\frac{E(R_r)-R_F}{\sigma}$. This is similar to the line on the efficient frontier plot generated by drawing a line tangent to the efficient frontier (point T) starting from an intercept point equal to the risk-free rate R_F.

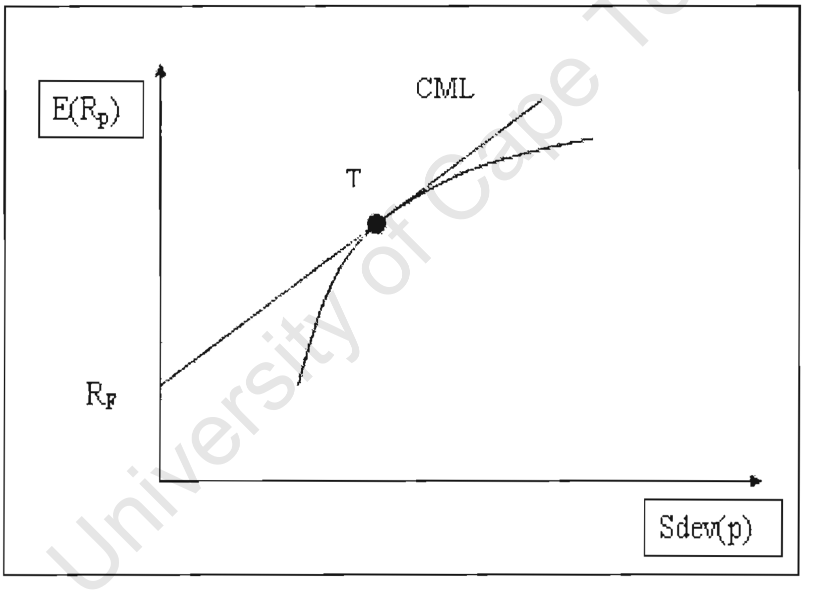

Figure 2.1: CML Line

At point T the investor is investing in a portfolio only consisting of securities, while at point R_F they are investing solely in the risk-free rate. In between points T and R_F (lending portfolio) the investor is investing part of their wealth in equities and the rest in a risk-free asset. Investing on line R_FT gives investors more options than just investing on the efficient frontier, for decreasing volatilies from point T line R_FT is higher than all other points on the frontier and therefore offers a higher return for the same volatility.

If we move to the right of point T (borrowing portfolio), then the investor invests all their wealth, plus an additional borrowed amount. The straight line passing through points R_F and T shows the range of portfolios an investor can invest in by lending and borrowing at the risk-free rate, and is called the capital market line (CML), since the portfolio at point T (on the efficient frontier) contains the optimal portfolio consisting only of equities, it is known as the market portfolio, so from equation (2.18) the CML is.

$$E(R_p) = R_F + \frac{E(R_T) - R_F}{\sigma_T} Sdev(p) \tag{2.19}$$

$E(R_p)$ is the expected rate of return and $Sdev(p)$ is the standard deviation of any portfolio on the CML, R_F is the risk-free rate, σ_T is the standard deviation of the market portfolio, $E(R_T)$ is the expected rate of return of the market portfolio, $\frac{E(R_T)-R_F}{\sigma_T}$ is called the market price of risk. From equation (2.19) we deduce that the expected rate of return for a portfolio on the CML consists of the risk-free rate plus a premium of $\frac{E(R_T)-R_F}{\sigma_T} Sdev(p)$.

CAPM

The CAPM model is concerned with the development of a security market line (SML) for a given security or portfolio of securities, which is the equation for the equilibrium expected return of the portfolio based on it's sensitivity to the market.

If we take the standard deviation as our measure of risk, then the systematic/market risk of an individual asset i is equal to.

$$\sigma_i^s = \beta_i \sigma_M$$

Since the market standard deviation σ_M is the same for all assets, we can consider β as the relative measure for the systematic risk of the assets. For a portfolio the systematic risk will be.

$$
\begin{aligned}
\sigma_p^s &= \beta_p \sigma_M \ where \\
\beta_p &= \sum_{i=1}^{n} w_i \beta_i \quad and \quad \sum_{i=1}^{n} w_i = 1
\end{aligned}
$$

So that the systematic risk of a portfolio is the weighted average of the systematic risks of the assets which make up the portfolio. Now we formulate a relationship between risk and return as follows.

If an investor holds a portfolio with $\beta_r = 1$ then the portfolio earns an expected rate of return equal to the market portfolio, but if $\beta_f = 1$ then the portfolio earns an expected rate of return equal to the risk-free rate R_F. If the investor invests a portion w in a risky portfolio with $\beta_r = 1$ and $(1 - w)$ in a risk-free portfolio with $\beta_f = 0$, then the beta of the composite portfolio β_p is.

$$
\begin{aligned}
\beta_p &= w\beta_r + (1 - w).\beta_f \\
&= w \cdot 1 + (1 - w) \cdot 0 \\
&= w
\end{aligned}
\tag{2.20}
$$

Equation (2.20) shows that the beta of this composite portfolio is equal to the wealth invested in the risky portfolio. If 100 % or less, of the investors wealth is invested in the risky portfolio then, $0 \leq \beta_p \leq 1$ (a lending portfolio), If the investor has a borrowing/leveraged portfolio then $\beta_p > 1$. The expected return of the composite portfolio is given by (2.15). Substituting for w from equation (2.20) we get.

$$
\begin{aligned}
E(R_p) &= \beta_p E(R_M) + (1 - \beta_p) R_F \\
E(R_p) &= R_F + \beta_p [E(R_M) - R_F]
\end{aligned}
\tag{2.21}
$$

The CAPM model, equation (2.21) states that there exists a linear relationship between the expected return of a portfolio, $E(R_p)$ and it's systematic risk, β_p. If β_p is 0 then the expected rate of return for the portfolio is equal to the risk free rate R_F, but if $\beta_p > 0$ then the investor receives a premium for accepting market risk equal to $\beta_p[E(R_M) - R_F)]$.

The security market line (SML) for an individual security is similar to equation (2.21) and is given by [18] as:

$$E(R_i) = R_F + \beta_i[E(R_M) - R_F] \tag{2.22}$$

The CAPM model allows us to estimate the expected return of a portfolio or securities given R_M, R_F and β.

Let us compare the CAPM model to the market model equation (2.12), the corresponding market model for a portfolio is.

$$R_p = \alpha_p + \beta_p R_M + U_p \tag{2.23}$$

If we restate equation (2.23) in risk premium form we get.

$$\begin{aligned} R_p - R_F &= [\alpha_p - R_F(1-\beta_p)] + \beta_p(R_M - R_F) + U_p \quad or \\ R_p &= [\alpha_p - R_F(1-\beta_p)] + R_F + \beta_p(R_M - R_F) + U_p \end{aligned} \tag{2.24}$$

$[\alpha_p - R_F(1-\beta_p)]$ may be interpreted as the portfolio risk premium, if the market risk premium is zero the market rate of return equals the risk-free rate. If we take the expectation of equation (2.24) we get.

$$E(R_p) = [\alpha_p - R_F(1-\beta_p)] + R_F + \beta_p(E(R_M) - R_F) \tag{2.25}$$

If we compare the above equation to the CAPM model equation (2.21).

$$E(R_p) = R_F + \beta_p[E(R_M) - R_F]$$

Then, if the first term in equation (2.25) (risk-adjusted excess return) is positive, $\alpha_p > R_F(1-\beta_p)$ then the expected return is greater then that predicted by the CAPM model and an investor holding such a portfolio tends to beat the market, If a portfolio has a negative risk-adjusted excess return then it tends to under perform compared to the market.

A few critique points of the CAPM model are [18]:

- Roll critique 1: If the market portfolio is not mean-variance efficient, it is impossible to accept or reject the presumed linear relationship between expected return and risk, so tests of the CAPM should be interpreted with great caution;

- Roll critique 2: Roll argues that using a proxy for the market portfolio as a benchmark is invalid if the proxy is not mean variance efficient (does not lie on the efficient frontier), then it is impossible to accept or reject the CAPM;

- Benchmark Error 1: This error occurs when the risk level of a portfolio β_p is calculated incorrectly due to the use of an inefficient market index. If the incorrect risk level is $\hat{\beta}_p$ and the correct risk level is β_p such that $\hat{\beta}_p \leq \beta_p$, and $E(\hat{R}_s) \leq R_p \leq E(R_s)$, where R_p is the realised return of the portfolio, then the investor thinks they have over performed relative to the market, while they have actually under performed;

- Benchmark Error 2: This error is due to the position of the SML line being incorrect, again due to the use of an inefficient market index. In this case there is either a difference between the true risk free rate and the one used in the model or a difference between the expected return $E(\hat{R_M})$ of the market index proxy and that of the mean-efficient market index $E(R_M)$.

In summary. Errors in the CAPM model seem to occur when the chosen market index proxy is not a mean-variance efficient portfolio.

Chapter 3

Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR)

3.1 Value at Risk (VaR)

Value at risk aims to provide portfolio managers with a single value to summarise the market risk of their portfolio. VaR tries to answer the question: " What is the maximum amount we expect to lose over N days with a certainty of P percent", where N is the period, in days, over which we expect the loss to occur and P is the ζ confidence interval. There are even published works [20], that suggest analysts can meaningfully compare the risk profiles of different banks utilizing their VaR disclosures.

Since VaR is not only measured in days we need to be able to scale our volatility accordingly. The daily volatility of asset prices can be defined as the standard deviation of the underlying asset's daily returns [19], given the daily volatility of an asset we can find the T day volatility using the following equation.

$$\sigma_T = \sigma_{day}\sqrt{T} \tag{3.1}$$

We can divide our portfolio VaR calculations into two models, parametric and non-parametric.

Parametric models are based on statistical parameters such as mean and standard deviation of risk factors [6], examples are; Asset-Normal VaR (no risk factors), Delta-Normal VaR (delta approximation) and Delta-Gamma Normal VaR (delta-gamma approximation).

The non-parametric models use simulation or historical models to evaluate VaR and are can be either full valuation or partial valuation models. The full valuation models create a number of scenarios for the portfolio risk factors and revalue the entire portfolio accordingly, examples are; Monte Carlo, Historical simulation and Stress Scenarios. The partial valuation models simulate the risk factors but do not fully revalue the portfolio, instead it makes use of delta-gamma approximations to calculate the new portfolio value.

The Asset-Normal VaR is probably the most basic model for calculating VaR, this model makes the assumption that the instruments in the portfolio are normally distributed [6]. The VaR formula, over the period T-t, is then given by [6].

$$VaR(t,T) = z_\zeta \sqrt{v'\Sigma v}\sqrt{T-t} \tag{3.2}$$

where z_ζ is the $\zeta - \mathit{quantile}$, the vector v contains the amount invested in each asset and Σ is the covariance matrix of daily volatilities.

Asset-Normal VaR Example: We have positions in 2 stocks A and B worth $R5$ million and $R1.5$ million respectively with daily volatilities of $\sigma_A = 2\%$, $\sigma_B = 1\%$ and a correlation coefficient $\rho = 0.7$. Let us calculate the 10 day VaR at the 99% confidence level.

$$\begin{aligned} VaR(t,T) &= 2.33\sqrt{\begin{pmatrix} 5 & 1.5 \end{pmatrix}\begin{pmatrix} 0.04\% & 0.01\% \\ 0.01\% & 0.01\% \end{pmatrix}\begin{pmatrix} 5 \\ 1.5 \end{pmatrix}}\sqrt{10} \\ &= 2.33 \cdot 0.111018 \cdot \sqrt{10} \\ &= R817,992.62 \end{aligned}$$

We can therefore say with 99% confidence that our portfolio will not lose more than R817,992.62 over a 10 day period.

By using the square root of time rule, we have made the assumption that the volatilities of the risk factors are constant over the period VaR is calculated and that there is no serial correlation. In practice this is however not the case as the time series of underlying risk factors show signs of mean reversion [6].

The Monte Carlo method makes use of an assumed probability distribution (e.g. log-normal for stock prices) of the risk factors to generate future scenarios for the time horizon for which we are calculating the ζVaR, the portfolio is then revalued for each scenario. The profit/loss is then calculated by subtracting the portfolio value for each scenario from the current portfolio value, sorting the n losses in ascending order the m^{th} largest loss, where m is the integer of the value $n\zeta$, is the VaR of the portfolio.

3.2 Why a Coherent measure of risk ?

We will discuss some of the properties of coherent risk measures, which will help to explain why they are a preferred set of risk measures when compared to the traditional VaR risk measure.

CVaR is a risk measure which falls into the category of coherent risk measures, [8] [2] [1] [25]. Let $\rho(X)$ be the function used to measure risk then a few axioms that need to be satisfied for a risk measure are [8]:

Definition 1: Consider a set W of real-valued random variables. A function $\rho : W \longrightarrow \mathbb{R}$ is called a risk measure if, for $X, Y \in W$ it satisfies the following properties.

- **Relevance**: If $X \leq 0$, but $X \neq 0$ then $\rho(X) \geq 0$. A portfolio with negative gains has a positive risk measure.

- **Monotonicity**: For Two different risks X and Y, where $X \geq Y$: $\rho(X) \leq \rho(Y)$.

In other words if one portfolio provides greater returns then the other for all states of the world it must be less risky.

- **Positive Homogeneity**: For every risk X and $\lambda \geq 0$:
  $\rho(\lambda X) = \lambda\rho(X)$.

- **sub-additivity**: For any two risks X and Y:
  $\rho(X + Y) \leq \rho(X) + \rho(Y)$.
  The sum of the risks of a portfolio must be an upper bound of the true total risk incurred by the portfolio.

- **Translation Invariance**: For every X and $\kappa \in \mathbb{R}$:
  $\rho(X + \kappa) = \rho(X) - \kappa$.
  If profits increase in all states of the world, then the risk must decrease in all states of the world.

A risk measure that satisfies the axioms of positive homogeneity, monotonicity, sub-additivity and translation invariance is said to be coherent, [8] [25], while CVaR satisfies all of these axioms VaR does not satisfy all of them. Sub-additivity is one of the axioms not satisfied by VaR, this means that the VaR of a portfolio of two components may be greater than the sum of the VaR of the individual components.

Example of the Non Sub-additivity of VaR

Let us Consider the probability space $(\Omega, A, \mathbb{P})$ with $\Omega = \{w_1, w_2, w_3\}$, where A is the set of all subsets of Ω and we specify $\mathbb{P}$ by:

$$\mathbb{P}[\{w_1\}] = \mathbb{P}[\{w_2\}] = p, \mathbb{P}[\{w_3\}] = 1 - 2p$$

Let us define $VaR_\zeta(X)$ as:

$$VaR_\zeta(X) = min\{x \in \mathbb{R} | P(X \leq x) \geq \zeta\} \tag{3.3}$$

If we then choose $0 < p < \frac{1}{3}$, $0 < \zeta < 2p$, fix some positive number N and let $X_i, i = 1, 2$ be two random variables defined on our probability space such that:

$$X_i(w_j) = \begin{cases} -N & \text{, if } i = j \\ 0 & \text{, otherwise} \end{cases}$$

If we choose $p = \frac{1}{5}$, $\zeta = \frac{1}{4}$ and $N = 5$, we get the following table:

	p	X_1	X_2	X_1+X_2
ω_1	1/5	-5	0	-5
ω_2	1/5	0	-5	-5
ω_3	3/5	0	0	0

Table 3.1: VaR

From the definition of VaR_ζ it is easy to see that:

$$VaR_\zeta(X_1) = VaR_\zeta(X_2) = 0 \tag{3.4}$$

$$VaR_\zeta(X_1 + X_2) = 5 \tag{3.5}$$

And we have shown the non-sub-additivity of VaR.

3.3 C-VaR Introduction

In this section we introduce CVaR, we will denote vectors with bold letters and scalars with normal font. CVaR, which is also referred to as the Mean shortfall or Tail VaR, is the expected loss exceeding VaR.

Let us consider a portfolio made up of a combination of n components, we use a decision vector $\mathbf{x} = (x_1, ..., x_n)$ which is an element of the set of all possible portfolios $\mathbf{X} \subset \mathbb{R}^n$ subject to $\sum x_i = 1$, to decide on the composition of the portfolio. Let $L = f(\mathbf{x}, \mathbf{y})$ be the loss associated with $\mathbf{x}$ and some random vector $\mathbf{y} \in \mathbb{R}^m$ so that $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$. $\mathbf{y}$ represents future uncertainties in the market and is governed by a probability measure $\mathbb{P}$ which is independent of $\mathbf{x}$, let $\mathbf{y}$ have a density in $\mathbb{R}^m$ of $p(\mathbf{y})$ [23]. Then the loss function has a distribution in $\mathbb{R}$ induced by that of $\mathbf{y}$, and the probability of $f(\mathbf{x},\mathbf{y})$ not exceeding a given limit of α is.

$$\begin{aligned}\vartheta(\mathbf{x},\alpha) &= \mathbf{P}\{\mathbf{y}|f(\mathbf{x},\mathbf{y}) \leq \alpha\} \\ &= \int_{f(\mathbf{x},\mathbf{y})\leq\alpha} p(\mathbf{y})dy \qquad (3.6)\end{aligned}$$

If we keep the decision vector $\mathbf{x}$ invariant and vary α in equation (3.6) we generate the distribution function for the loss function $f(\mathbf{x},\mathbf{y})$. For simplicity to calculate CVaR, we will assume that this distribution function is continuous everywhere with respect to α. The ζVaR is then defined as:

$$\alpha_\zeta(\mathbf{x}) = min\{\alpha \in \mathbb{R} : \vartheta(\mathbf{x},\alpha) \geq \zeta\} \qquad (3.7)$$

Let us assume that the future uncertainties in vector $\mathbf{y}$ all occur with the same probability, so for a sample of size J the sample points $\mathbf{y}_i$, i =1,...,J will each have probability 1/J. This allows our losses L, for each decision vector $\mathbf{x}$ to also be discretely distributed for a given sample J. If we then order these losses $L_j = f(\mathbf{x},y_i), i = 1,...,J$ in ascending order $L_1 \leq L_2 \leq ... \leq L_J$ then ζVaR=L_k, where $k = J\zeta$ if $J\zeta = integer(J\zeta)$ else $k = J\zeta + 1$. For example if we have J = 100 and $\zeta = 0.9$, ζVaR=L_{90}. In order to minimise VaR we would be interested in the following problem.

$$\min_{x\in X\subset\mathbf{R}^n} \alpha_\zeta(\mathbf{x})$$

Due to the difficulty in minimising VaR [21], because it has multiple minima, we choose to minimise CVaR instead. CVaR is the expected value of the losses exceeding VaR, so for a given ζ.

$$\begin{aligned}\zeta CVaR &= \Phi_\zeta(\mathbf{x}) = \mathbb{E}[f(\mathbf{x},\mathbf{y})|f(\mathbf{x},\mathbf{y}) \geq \alpha_\zeta(\mathbf{x})] \\ &= (1-\zeta)^{-1}\int_{f(\mathbf{x},\mathbf{y})\geq\alpha_\zeta(\mathbf{x})} f(\mathbf{x},\mathbf{y})p(\mathbf{y})dy \qquad (3.8)\end{aligned}$$

For discrete distributions CVaR is defined as the weighted average of VaR plus the losses strictly exceeding VaR, [21]. If we denote the conditional expectation of losses strictly exceeding VaR as $\Phi_\zeta(\mathbf{x})^+$, then

$$\Phi_\zeta(\mathbf{x}) = \delta\alpha_\zeta(\mathbf{x}) + (1-\delta)\Phi_\zeta(\mathbf{x})^+ \tag{3.9}$$
$$where \quad \delta = [\vartheta(\mathbf{x}, \alpha_\zeta(\mathbf{x})) - \zeta]/[1-\zeta] \in [0,1]$$

We can characterise the VaR and CVaR functions on $X \times \mathbb{R}$ using a function F_ζ, [23].

$$F_\zeta(\mathbf{x}, \alpha) = \alpha + (1-\zeta)^{-1} \int_{y \in \mathbf{R}^m} [f(\mathbf{x}, \mathbf{y}) - \alpha]^+ p(\mathbf{y}) dy \tag{3.10}$$

Theorem 1 [1] *If $F_\zeta(x, \alpha)$ is continuously differentiable and convex as a function of α, the ζCVaR of the loss associated with any $x \in X$ can be determined from.*

$$\Phi_\zeta(x) = \min_{\alpha \in \mathbf{R}} F_\zeta(x, \alpha) \tag{3.11}$$

Let the set consisting of the values of α for which the minimum is obtained for the above equation be.

$$A_\zeta(x) = arg \min_{\alpha \in \mathbf{R}} F_\zeta(x, \alpha) \tag{3.12}$$

Then the above set is a nonempty, closed, bounded interval, and the ζVaR (which is the minimum alpha) of the loss is given by.

$$\alpha_\zeta(x) = leftendpoint \quad of \quad A_\zeta(x) \tag{3.13}$$

In summary we always have:

$$\alpha_\zeta(x) \in arg \min_{\alpha \in \mathbf{R}} F_\zeta(x.\alpha) \quad and \quad \Phi_\zeta(x) = F_\zeta(x, \alpha_\zeta(x)) \tag{3.14}$$

Where the $arg \min$ expression means that we want the value of the given argument $A_\zeta(\mathbf{x})$ for which the value of the given expression $F_\zeta(\mathbf{x}, \alpha)$ attains it's minimum.

The above theorem is important because of the ease with which continuously differentiable convex functions can be minimised numerically. As mentioned before minimising ζVaR can be a very complicated matter, however we can

[1] See Appendix for Proof

minimise the convex function ζCVaR, which is always greater than ζVaR, and obtain ζVaR as a byproduct of this minimization. ζVaR will be the minimum value of α after minimising $F_\zeta(\mathbf{x}, \alpha)$.

We made the assumption earlier that the future scenarios in vector $\mathbf{y}$ each have an equal probability of occurrence, so for a sample of J scenarios the probability of each y_i occurring is $d\mathbb{P} = 1/J$. With this assumption we can approximate the integral in $F_\zeta(\mathbf{x}, \alpha)$ as follows.

$$\begin{aligned} & \int_{y\in\mathbf{R}^m} [f(\mathbf{x},\mathbf{y}) - \alpha]^+ p(\mathbf{y})dy \\ = & \int_{y\in\mathbf{R}^m} [f(\mathbf{x},\mathbf{y}) - \alpha]^+ d\mathbb{P} \\ = & \; J^{-1} \sum_{k=1}^{J} [f(\mathbf{x},\mathbf{y}_k) - \alpha]^+ \end{aligned} \tag{3.15}$$

Our new linear function with respect to α then becomes.

$$\hat{F}_\zeta(\mathbf{x},\alpha) = \alpha + \frac{1}{J(1-\zeta)} \sum_{k=1}^{J} [f(\mathbf{x},\mathbf{y}_k) - \alpha]^+ \tag{3.16}$$

The CVaR optimization problem for J scenarios and n securities can now be formulated as.

$$\begin{aligned} \min_{x,\alpha} \quad & \hat{F}_\zeta(\mathbf{x},\alpha) \\ \textit{subject} \quad & \textit{to} \\ \sum_{i=1}^{n} w_i \quad = \quad & 1, 0 \le w_i \le 1, i = 1, ..., n \end{aligned} \tag{3.17}$$

The resulting α is the VaR of the portfolio and the $x - vector$ the optimal portfolio.

Let us make sure that our CVaR function (3.16) is indeed a coherent measure of risk, but first let us introduce some dominance relationships [25]. Let X and Y be two random variables.

- *Stochastic dominance of order 1*: The relation

$$X \prec_{SD(1)} Y$$

holds iff

$$\mathbf{E}[\psi(X)] \leq \mathbf{E}[\psi(Y)]$$

for all (integrable) monotonic functions ψ.

- Monotonic dominance of order 1: The relation

$$X \prec_{MD(1)} Y$$

holds iff

$$\mathbf{E}[\psi(X)] \leq \mathbf{E}[\psi(Y)]$$

for all (integrable) concave functions ψ.

- Stochastic dominance of order 2: The relation

$$X \prec_{SD(2)} Y$$

holds if and only if

$$\mathbf{E}[\psi(X)] \leq \mathbf{E}[\psi(Y)]$$

for all (integrable) concave, monotonic functions ψ.

So $X \prec_{SD(2)} Y$ holds iff $X \prec_{SD(1)} Y$ and $X \prec_{MD(1)} Y$.

Therefore $X \prec_{SD(2)} Y$ is equivalent to $\int_{-\infty}^{t} F_X(u)du \leq \int_{-\infty}^{t} F_Y(u)du \quad \forall t$.

Let's give a short proof of the statement above. Since $\int_{-\infty}^{t} F(u)du = \int_{-\infty}^{\infty}[t-u]^+ dF(u)$, we see that $X \prec_{SD(2)} Y$ is similar to $\int_{-\infty}^{\infty} \psi(u)dF_X(u) \leq \int_{-\infty}^{\infty} \psi(u)dF_Y(u)$ for all functions $\psi(u) = \sum_k(-\alpha_k)[t_k - u]^+ + \beta_k, \quad with \quad \alpha_k \geq 0$, these functions are contained in the set of all concave, monotone functions.

Proposition. C-VaR has the following properties:

1. The function $y \to [y-\alpha]^+$ is convex, so if X and Y are two random variables and $[t]^+ = t$ if $t > 0$ else $t = 0$. Then for $0 < \lambda < 1$ we have the following.

$$[\lambda(X-\alpha) + (1-\lambda)(Y-\alpha)]^+ \leq \lambda[X-\alpha]^+ + (1-\lambda)[Y-\alpha]^+$$

2. CVaR is *translation invariant* i.e.

$$CVaR(X + c) = CVaR(X) + c$$

3. CVaR is *positively homogeneous* i.e.

$$CVaR(cX) = cCVaR(X)$$

4. CVaR is *convex* i.e. for $0 < \lambda < 1$

$$CVaR(\lambda X + (1 - \lambda)Y) \leq \lambda CVaR(X) + (1 - \lambda)CVaR(Y)$$

5. CVaR is *monotonic* w.r.t. SD(2) and therefore SD(1) i.e.

$$X \prec_{SD(2)} Y \quad then$$
$$CVaR(X) \leq CVaR(Y)$$

6. CVaR is *monotonic* w.r.t. MD(2) i.e.

$$X \prec_{MD(2)} Y \quad then$$
$$CVaR(X) \leq CVaR(Y)$$

Proof. Let us prove (1.) if X and Y are two random variables, $[t]^+ = t$ if $t > 0$ else $t = 0$ and for $0 < \lambda < 1$ we will prove the inequality, since the equality is evident in the cases when $\alpha > XY$ and $\alpha < XY$.

1. Let $Y > \alpha > X$, we want.

$$
\begin{aligned}
& [\lambda(X - \alpha) + (1 - \lambda)(Y - \alpha)]^+ < \lambda[X - \alpha]^+ + (1 - \lambda)[Y - \alpha]^+ \\
\Rightarrow \quad & [\lambda X - \lambda\alpha + Y - \alpha - \lambda Y + \lambda\alpha)]^+ < Y - \alpha - \lambda Y + \lambda\alpha \\
\Rightarrow \quad & [\lambda(X - Y) + Y - \alpha]^+ < Y - \alpha + \lambda(\alpha - Y) \\
\Rightarrow \quad & [Y - \alpha - \lambda(Y - X)]^+ < (Y - \alpha) - \lambda(Y \quad \alpha)
\end{aligned}
$$

The right hand side of this inequality is always greater than zero. For the left hand side we have two possibilities, first $Y - \alpha \leq \lambda(Y - X)$ in which case we have.

$$0 < (Y - \alpha) - \lambda(\alpha - Y) \geq 0$$

For the second case $Y - \alpha > \lambda(Y - X)$ and we get.

$$
\begin{aligned}
& (Y - \alpha) - \lambda(Y - X) < (Y - \alpha) - \lambda(Y - \alpha) \\
\Rightarrow \quad & -\lambda(Y - X) < -\lambda(Y - \alpha)
\end{aligned}
\tag{3.18}
$$

Which is what we want to show. From the definition of CVaR (2.) and (3.) are obvious so lets prove (4.).

Let α_i be such that $CVaR(X) = \alpha_i + \frac{1}{1-\zeta}\mathbf{E}[X - \alpha]^+$, and since $y \rightarrow [y - \alpha]^+$ is convex, we have.

$$
\begin{aligned}
& CVaR(\lambda X + (1 - \lambda)Y) \\
= \quad & \lambda\alpha_1 + (1 - \lambda)\alpha_2 + \frac{1}{1-\zeta}\mathbb{E}[\lambda X + (1 - \lambda)Y - (\lambda\alpha_1 + (1 - \lambda)\alpha_2)]^+ \\
\leq \quad & \lambda\alpha_1 + (1 - \lambda)\alpha_2 + \frac{\lambda}{1-\zeta}\mathbb{E}[X - \alpha_1]^+ + \frac{1-\lambda}{1-\zeta}\mathbb{E}[Y - \alpha_2]^+ \\
= \quad & \lambda CVaR(X) + (1 - \lambda)CVaR(Y)
\end{aligned}
$$

Properties (5.) and (6.) follow from our dominance relations and the fact that $y \rightarrow [y - \alpha]^+$ is monotone and convex. CVaR therefore satisfies the requirements for a coherent risk measure.

Chapter 4

Dependence concepts In Financial Risk Management

In this chapter we will discuss the concepts commonly used in financial risk management to describe the dependence structure between the securities in a portfolio. We start with linear correlation following the discussion of [17].

4.1 Linear Correlation

For two real-valued random variables X and Y with finite variances $\sigma^2[X]$ and $\sigma^2[Y]$ respectively, the covariance between them is defined as.

$$Cov[X,Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \tag{4.1}$$

Definition 2: The linear correlation coefficient between X and Y is.

$$\rho(X,Y) := \frac{Cov[X,Y]}{\sqrt{\sigma^2[X]\sigma^2[Y]}} \tag{4.2}$$

The linear correlation $\rho(X,Y)$ is a measure of linear dependence between the random variables. For perfect linear dependence ($Y = aX + b$) with $a \in \mathbb{R}\backslash\{0\}$ and $b \in \mathbb{R}$ we have $\rho(X,Y) = \pm 1$. We can show this by a representation of the change in variance of Y with respect to the regression on X.

$$\rho(X,Y)^2 = \frac{\sigma^2[Y] - \min_{a,b} \mathbb{E}[(Y - (aX + b))^2]}{\sigma^2[Y]} \tag{4.3}$$

Equation (4.3) equals 1 in the case of perfect linear regression which is why we have the value of ± 1 for linear correlation in the case of perfect dependence. The constants a and b which minimise the squared difference, $\mathbb{E}[(Y-(aX+b))^2]$ in equation (4.3) for a perfect linear correlation are[1].

$$\begin{aligned} a &= \frac{Cov[X,Y]}{\sigma^2[X]} \\ b &= \mathbb{E}[Y] - a\mathbb{E}[X] \end{aligned}$$

In the case of independent random variables $\rho(X,Y) = 0$ since $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ so Cov[X,Y] = 0, but zero correlation does not (necessarily) imply independence. For imperfect linear dependence we have -1< $\rho(X,Y)$ <1, we are interested in investigating this type of linear dependence.

An important property of linear correlation is that it is constant under strictly increasing (positive) linear transformations[2], so

$$\rho(\alpha X + \beta, \gamma Y + \delta) = sgn(\alpha \cdot \gamma)\rho(X,Y)$$

where $\alpha, \gamma \in \mathbb{R}\backslash\{0\}$ and $\beta, \delta \in \mathbb{R}$. If we have two vectors $\mathbf{X}=(X_1, ..., X_n)^t$ and $\mathbf{Y}=(Y_1, ..., Y_n)^t$, and two affine linear transformations A:$\mathbb{R}^n \rightarrow \mathbb{R}^m, x \mapsto Ax + a$ and B:$\mathbb{R}^n \rightarrow \mathbb{R}^m, x \mapsto Bx + b$ for $A, B \in \mathbb{R}^{m \times n}, a, b \in \mathbb{R}^m$, then we have.

$$Cov[A\mathbf{X} + a, B\mathbf{Y} + b] = ACov[\mathbf{X},\mathbf{Y}]B^t$$

If we are interested in one random vector $\mathbf{X}$ then we let $\mathbf{Y} = \mathbf{X}$ and $\rho(\mathbf{X}) := \rho(\mathbf{X},\mathbf{X})$ and Cov[$\mathbf{X}$] := Cov[$\mathbf{X}$,$\mathbf{X}$], then for every linear combination[3] $\alpha^t\mathbf{X}$, $\alpha \in \mathbb{R}^n$ of the elements of vector $\mathbf{X}$ we have a relationship between the covariance and variance such that.

$$\sigma^2[\alpha^t\mathbf{X}] = \alpha^t Cov[X]\alpha \tag{4.4}$$

[1] proof in Appendix
[2] proof in Appendix
[3] important for portfolio theory

4.1.1 A few undesirable properties of Linear Correlation

Linear correlation has difficulty with heavy-tailed distributions, because when the variances of X and Y are not finite, linear correlation is not defined.

Linear correlation is also not constant under non-linear strictly increasing transformations [17] K:$\mathbb{R} \to \mathbb{R}$, so we have.

$\rho(K(X),(K(Y)) \neq \rho(X,Y)$

We mentioned before that independence of two random variables means they are uncorrelated and the coefficient of correlation $\rho = 0$, but a coefficient of zero correlation does not necessarily imply independence among two random variables. An example of this is if we take $X \sim N(0,1)$ and $Y = X^2$, since the third moment of the standard normal distribution is zero the covariance between X and Y is zero and the correlation coefficient is also zero. [17] Mentions that we should only interpret uncorrelatedness as implying independence for multivariate distributions where both the marginal and joint distributions are normal.

4.2 Conditional Multivariate Normal Distribution

In this section we show how to arrive at the conditional mean and covariance matrix used for multivariate normal random number generation, we follow the discussion of [11]. Let $\mathbf{X}$ have a multivariate normal distribution $\mathbf{X} \sim \mathrm{N}(\mu, \Sigma)$ with s elements and let $\mathbf{X}$ be partitioned as.

$$\mathbf{X} = \begin{pmatrix} X^{(1)} \\ X^{(2)} \end{pmatrix}$$

Where $\mathbf{X}^{(1)}$ has q elements and $\mathbf{X}^{(2)}$ has $r = s - q$ elements. If we now partition μ and Σ such that

$$\mu = \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix} \quad and \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

and perform a nonsingular transformation.

$$\mathbf{Y}^{(1)} = \mathbf{X}^{(1)} + \mathbf{M}\mathbf{X}^{(2)} \tag{4.5}$$

$$\mathbf{Y}^{(2)} = \mathbf{X}^{(2)} \tag{4.6}$$

Now if we choose $\mathbf{M}$ so that the components of $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ are uncorrelated, the matrix $\mathbf{M}$ must satisfy the following equation.

$$\begin{aligned} 0 &= E(\mathbf{Y}^{(1)} - E\mathbf{Y}^{(1)})(\mathbf{Y}^{(2)} - E\mathbf{Y}^{(2)})' \\ &= E(\mathbf{X}^{(1)} + \mathbf{M}\mathbf{X}^{(2)} - E\mathbf{X}^{(1)} - \mathbf{M}E\mathbf{X}^{(2)})(\mathbf{X}^{(2)} - E\mathbf{X}^{(2)})' \\ &= E(\mathbf{X}^{(1)} - E\mathbf{X}^{(1)} + \mathbf{M}(\mathbf{X}^{(2)} - E\mathbf{X}^{(2)}))(\mathbf{X}^{(2)} - E\mathbf{X}^{(2)})' \\ &= \Sigma_{12} + \mathbf{M}\Sigma_{22} \end{aligned} \tag{4.7}$$

so

$$\mathbf{M} = -\Sigma_{12}\Sigma_{22}^{-1} \tag{4.8}$$

$$\mathbf{Y}^{(1)} = \mathbf{X}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{X}^{(2)} \tag{4.9}$$

The vector $\mathbf{Y} = \mathbf{C}\mathbf{X}$ is a nonsingular linear transformation of $\mathbf{X}$ where.

$$\mathbf{C} = \begin{pmatrix} \mathbf{I} & -\Sigma_{12}\Sigma_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

$\mathbf{Y}$ has distribution $\mathbf{Y} \sim \mathrm{N}(\mathbf{C}\mu, \mathbf{C}\Sigma\mathbf{C}')$ so that

$$\begin{aligned} \mathbf{C}\mu &= \begin{pmatrix} \mathbf{I} & -\Sigma_{12}\Sigma_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \cdot \begin{pmatrix} \mu^{(1)} \\ \mu^{(2)} \end{pmatrix} \\ &= \begin{pmatrix} \mu^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)} \\ \mu^{(2)} \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned} \mathbf{C}\Sigma\mathbf{C}' &= \begin{pmatrix} \mathbf{I} & -\Sigma_{12}\Sigma_{22}^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \cdot \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ -\Sigma_{22}^{-1}\Sigma_{21} & \mathbf{I} \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22} \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_{11.2} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22} \end{pmatrix} \end{aligned}$$

where $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ are therefore independently distributed with the following marginal distributions.

$$\begin{aligned}\mathbf{Y}^{(1)} &\sim N(\mu^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)}, \Sigma_{11.2})\\ \mathbf{Y}^{(2)} &\sim N(\mu^{(2)}, \Sigma_{22})\end{aligned}$$

If we use the transformations in (4.5) to transform back from $\mathbf{Y}$ to $\mathbf{X}$, noting that the Jacobian is 1 the joint density of $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$, $f(x^{(1)}, x^{(2)})$ is

$$\begin{aligned}f(x^{(1)}, x^{(2)}) &= f(\mathbf{y}^{(1)}/\mu^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)}, \Sigma_{11.2}) \cdot f(\mathbf{y}^{(2)}/\mu^{(2)}, \Sigma_{22}) \cdot |J|\\ &= (2\pi)^{-q/2}|\Sigma_{11.2}|^{-1/2}\\ &\times\ exp((-1/2)(\mathbf{x}^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{x}^{(2)} - (\mu^{(1)} - \Sigma_{12}\Sigma_{22}^{-1}\mu^{(2)}))'\Sigma_{11.2}^{-1}\\ &\times\ (2\pi)^{-(s-q)/2}|\Sigma_{22}|^{-1/2}exp((-1/2)(\mathbf{x}^{(2)} - \mu^{(2)})'\Sigma_{22}^{-1}\\ &= f(\mathbf{x}^{(1)}/\mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}^{(2)} - \mu^{(2)}), \Sigma_{11.2}) \cdot f(\mathbf{x}^{(2)}/\mu^{(2)}, \Sigma_{22})\\ &= f(\mathbf{x}/\mu, \Sigma)\end{aligned}$$

since $f(x_1/x_2) = \frac{f(x_1,x_2)}{f(x_2)}$, the conditional distribution of $\mathbf{X}^{(1)}$ given $\mathbf{X}^{(2)} = \mathbf{x}^{(2)}$ becomes.

$$\begin{aligned}f(x_1/x_2) &= \frac{f(\mathbf{x}/\mu, \Sigma)}{f(\mathbf{x}^{(2)}/\mu^{(2)}, \Sigma_{22})}\\ &= f(\mathbf{x}^{(1)}/\mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}^{(2)} - \mu^{(2)}), \Sigma_{11.2})\end{aligned}$$

The above is a multivariate normal density with a conditional mean and covariance matrix of.

$$E(x_1/x_2) = \mu^{(1)} + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}^{(2)} - \mu^{(2)}) \tag{4.10}$$

$$cov(x_1/x_2) = \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \tag{4.11}$$

Note that the conditional mean depends on $\mathbf{x}_2$ while the conditional covariance matrix does not. The importance of the above result, used in modern portfolio theory, is that the dependence structure between the securities in a portfolio is described by the correlation matrix if we assume the marginal and the multivariate distributions are normally distributed [17].

4.3 Copula

In order to utilise the loss function for the portfolio, future returns need to be forecasted. The usual assumptions are that the returns of the instruments in the portfolio are normally distributed and linear correlation is used as the measure of dependence between the financial instruments, these assumptions are then used to forecast the future portfolio returns.

These assumptions are not entirely correct as returns for instruments in portfolios usually have 'fat tails' and tail dependence, meaning extreme events happen more often than what is modeled using a normal distribution. An extreme event for one security could also lead to an extreme event for another security in the portfolio - extreme events appear in clusters [22]- this invalidates the assumption of a portfolio with a multivariate normal distribution consisting of instruments with normal marginal distributions [17].

A copula is the distribution function of a random vector in $\Re^n$ with uniform-(0,1) marginals. Copula can be used to construct multivariate distributions which are consistent with given marginal distributions and dependency parameters which are necessary to simulate dependent random vectors.

Proposition A

Let Z = F(X), then Z has a uniform distribution on [0,1].

Proof: $P(Z \leq z) = P(F(X) \leq z) = P(X \leq F^{-1}(z)) = F(F^{-1}(z)) = z$ which is the uniform cumulative distribution function.

Proposition B

Let U be uniform on [0,1] and let X = F^{-1}(U). Then the c.d.f. of X is F.

Proof: $P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$.

Proposition B can be used to generate random variables for continuous distributions from the uniform distribution. If we have random variables $X_1, ..., X_n$ then the dependence between the random variables is described by their joint distribution function.

$$F(x_1, ..., x_n) = \mathbb{P}[X_1 \leq x_1, ..., X_n \leq x_n].$$

Now assuming that each random variable $X_1, ..., X_n$ has a continuous marginal distribution $F_1, ..., F_n$, we use Proposition A to transform the components of the random vector $\mathbf{X}=(X_1, ..., X_n)^t$ to have standard Uniform, U(0,1), marginal distributions. The joint distribution function C of $(F_1(X_1), ..., F_n(X_n))^t$ is called the copula of the random vector $(X_1, ..., X_n)^t$ or the multivariate distribution F [17].

Theorem 2 (Sklar's theorem) *Given an n-dimensional distribution function F with continuous marginal distributions $F_1, ..., F_n$, there exists a unique n-copula $C : [0,1]^n \to [0,1]$ such that.*

$$\begin{aligned} F(x_1, ..., x_n) &= \mathbb{P}[F_1(X_1) \leq F_1(x_1), ..., F_n(X_n) \leq F_n(x_n)] \\ &= C(F_1(x_1), ..., F_n(x_n)). \end{aligned} \tag{4.12}$$

So, if we have a multivariate distribution F, with $F_1, ..., F_n$ being the marginal distributions then the function.

$$C(u_1, ..., u_n) = F(F_1^{-1}(u_1), ..., F_n^{-1}(u_n)) \tag{4.13}$$

is an n-copula. The above mentioned copula represents the multivariate dependence structure of $(X_1, ..., X_n)^t$ which links the univariate uniform marginals of the X_i's, another definition for a copula is therefore a multivariate distribution function defined on the unit cube $[0,1]^n$ having uniform marginal distributions [22]. If all the F_i are continuous then the copula is unique, whereas in the discrete case there will be more than one copula [17] [7]. The density function for the copula $C(u_1, ..., u_n)$ is.

$$c(u_1, ..., u_n) = \frac{\partial C(u_1, ..., u_n)}{\partial u_1, ..., \partial u_n} \tag{4.14}$$

The basic properties every copula function should have are as follows:

Definition 3: A Copula is any function $C : [0,1]^n \rightarrow [0,1]$ with the following properties:

- $C(x_1, ..., x_i, ..., x_n)$ is increasing in each x_i, meaning if $x_{i1} \leq x_{i2}$ then $C(x_1, ..., x_{i1}, ..., x_n) \leq C(x_1, ..., x_{i2}, ..., x_n)$;
- $C(x_1, ..., x_i, ..., x_n)$ is grounded, $C(x_1, ..., x_i, ..., x_n) = 0$ if at least one $x_i = 0$;
- $C(1, ..., 1, x_i, 1, ..., 1) = x_i$ for all $i \in \{1, ..., n\}, x_i \in [0,1]$;
- For all $(a_1, ..., a_n), (b_1, ..., b_n) \in [0,1]$ where $a_i \leq b_i$ we have $\sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1+\ldots+i_n} C(x_{1i_1}, ..., x_{ni_n}) \geq 0$ where $x_{j1} = a_j$ and $x_{j2=b_j}$ for all $j \in \{1, ..., n\}$.

If we want to work with log-returns of the portfolio instruments instead of simple returns we need to ensure the copula remains invariant when we make this transformation.

Proposition C

If random vector $\mathbf{X}=(X_1, ..., X_n)^t$ has a copula C, let $K_1, ..., K_n$ be increasing continuous[4] functions, then $(K_1(X_1), ..., K_n(X_n))^t$ will also have the same copula C.

Proof: Let $(U_1, ..., U_n)^t$ have distribution function C with continuous marginal distributions F_{X_i} so that $U_i = F_{X_i}(X_i)$. Then

$$C(F_{K_1(X_1)}(x_1), ..., F_{K_n(X_n)}(x_n))$$

$$= \mathbb{P}[U_1 \leq F_{K_1(X_1)}(x_1), ..., U_n \leq F_{K_n(X_n)}(x_n)]$$

[4]since the K_i are continuous $F^{-1}_{K_i(X_i)} = K_1 \circ F^{-1}_{X_1}$

$$= \mathbb{P}[F^{-1}_{K_1(X_1)}(U_1) \leq x_1, ..., F^{-1}_{K_n(X_n)}(U_n) \leq x_n]$$

$$= \mathbb{P}[K_1 \circ F^{-1}_{X_1}(U_1) \leq x_1, ..., K_n \circ F^{-1}_{X_n}(U_n) \leq x_n]$$

$$= \mathbb{P}[K_1(X_1) \leq x_1, ..., K_n(X_n) \leq x_n]$$

We see that the copula remains the same only the marginal distributions change. If the marginal distributions $F_1, ..., F_n$ have marginal density functions $f_1, ..., f_n$ then the multivariate density function $f(\mathbf{X})$ for the multivariate distribution $\mathrm{F}(\mathbf{X})$ is equal to.

$$f(\mathbf{X}) = c(u_1, ..., u_n) \prod_{i=1}^{N} f_i(x_i) \tag{4.15}$$

One of the most important advantages of equation (4.15) is that we can separate our modeling problem into two parts; the first deals with choosing the correct dependency structure and the second deals with the identification of the marginal distributions.

4.4 Summary of a few Copulas

[7] and [13] give a summary of a few copula distributions used in finance.

Copula of independent Marginals

$C\perp(u_1, u_2) = u_1 \cdot u_2$

Gaussian Copula

Let $\mathbf{R}$ be a symmetric, positive definite matrix with diag $\mathbf{R} = 1$ and $\Phi_{\mathbf{R}}$ the standard multivariate normal distribution, then the multivariate gaussian copula is:

$$C^{Ga}_{\mathbf{R}}(u_1, ..., u_N; \mathbf{R}) = \Phi_{\mathbf{R}}(\varphi^{-1}(u_1), ..., \varphi^{-1}(u_N)) \tag{4.16}$$

From equation (4.15) we can find the density from the multinormal distribution:

$$\begin{aligned}
\frac{1}{(2\pi)^{\frac{N}{2}}|R|^{\frac{1}{2}}} exp(-\frac{1}{2}\mathbf{x}^\top \mathbf{R}^{-1}\mathbf{x}) &= c_{\mathbf{R}}^{Ga}(\phi(u_1), ..., \phi(u_N)) \left(\prod_{n=1}^{N} \frac{1}{\sqrt{2\pi}} exp(-\frac{1}{2}x_n^2) \right) \\
\Rightarrow c_{\mathbf{R}}^{Ga}(u_1, ..., u_N) &= \frac{\frac{1}{(2\pi)^{\frac{N}{2}}|R|^{\frac{1}{2}}} exp(-\frac{1}{2}\varsigma^\top \mathbf{R}^{-1}\varsigma)}{\frac{1}{(2\pi)^{\frac{N}{2}}} exp(-\frac{1}{2}\varsigma^\top \varsigma)} \qquad (4.17) \\
&= \frac{1}{|R|^{\frac{1}{2}}} exp\left(-\frac{1}{2}\varsigma^\top (\mathbf{R}^{-1} - \mathbb{I})\varsigma\right) \qquad (4.18)
\end{aligned}$$

Where $x_1, ..., x_N$ are observations from the N instruments, $\varsigma_n = \phi^{-1}(u_n)$ and $\mathbb{I}$ is an identity matrix.

Students t-Copula

Let $\mathbf{R}$ be a symmetric, positive definite matrix with diag $\mathbf{R} = \mathbf{1}$ and $T_{\mathbf{R},\nu}$ the standard multivariate Student's t-distribution with ν degrees of freedom, then the multivariate Student's t-copula is:

$$C_{\mathbf{R}}^{T}(u_1, ..., u_N; \mathbf{R}) = T_{\mathbf{R},\nu}(t_\nu^{-1}(u_1), ..., t_\nu^{-1}(u_N)) \qquad (4.19)$$

Where t_ν^{-1} is the inverse univariate Student's t distribution, using equation (4.15) again the density for this copula is:

$$c_{\mathbf{R}}^{T}(u_1, ..., u_N; \mathbf{R}) = |R|^{-\frac{1}{2}} \frac{\Gamma(\frac{\nu+N}{2})[\Gamma(\frac{\nu}{2})]^N}{[\Gamma(\frac{\nu+1}{2})]^N \Gamma(\frac{\nu}{2})} \frac{(1+\frac{1}{\nu}\varsigma^\top \mathbf{R}^{-1}\varsigma)^{-\frac{\nu+N}{2}}}{\prod_{n=1}^{N}(1+\frac{\varsigma_n^2}{\nu})^{-\frac{\nu+1}{2}}} \qquad (4.20)$$

Gumbel Copula

The Gumbel copula is an extreme value copula.

$$C_{Gu}^{\delta}(u_1, u_2) = exp\left[-\{\tilde{u}_1^\delta + \tilde{u}_2^\delta\}^{\frac{1}{\delta}}\right] \qquad (4.21)$$

Where $\tilde{u} = -\log u$ and the parameter $\delta \geq 1$ gives the degree of dependency between the two random variables X_1 and X_2, if $\delta = 1$ then they are independent and as $\delta \to \infty$ they approach perfect dependency.

Chapter 5

Modeling Events in Practice

5.1 Parameter Estimation

The previous sections discuss statistical models we can use to describe the various marginal and multivariate distributions for our investigation. We have to remember that statistical models have to be mathematically tractable and are not necessarily a means to an end but a means to providing statistical inference. From an industrial point of view we have to remember that for high dimensions, if the parameter estimation and the simulation cannot be solved easily then the model is not tractable.

For our investigative purposes we will use the maximum likelihood estimation (MLE) method for parameter estimation of the marginal and multivariate distributions. But let's take a look at another popular method first, namely method of moments estimation.

5.1.1 Method of Moments Estimator

Let $f(x/\theta_1, ..., \theta_k)$ be a density function of $k - parameters, \theta_1, ..., \theta_k$ and let $\mu_1, ..., \mu_k$ be the moments about the origin so that.

$$\begin{aligned} \mu_r^{'} &= E(X^r) \\ &= \int x^r f(x/\theta_1, ..., \theta_k)dx \end{aligned} \tag{5.1}$$

Next let $(X_1, ..., X_n)$ be a random sample of $f(x/\theta_1, ..., \theta_k)$, then the sample moments will be

$$m_r^{'} = \frac{1}{n}\sum_{i=1}^{n} x_i^r \tag{5.2}$$

If we let $\tilde{\theta}_1, ..., \tilde{\theta}_k$ be the solutions to

$$m_r^{'} = \mu_r^{'}(\theta_1, ..., \theta_k) \tag{5.3}$$

Then these estimates $(\tilde{\theta}_1, ..., \tilde{\theta}_k)$ are called the method of moments estimators.

Example 5.1 Let $\mathbf{X} \sim N(\mu, \sigma^2)$

$$\mu_1^{'} = E(X) = \mu$$
$$\mu_2^{'} = E(X^2) = \mu^2 + \sigma^2$$

And we also have

$$m_1^{'} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x} = m_1$$
$$m_2^{'} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

Thus

$$\begin{aligned} m_1^{'} &= \mu_1^{'} = \mu \quad or \quad \hat{\mu} = \overline{X} = m_1 \\ m_2^{'} &= \mu^2 + \sigma^2 \quad or \\ \sigma^2 &= m_2^{'} - (m_1^{'})^2 \\ &= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \end{aligned}$$

So the method of moments estimators of μ and σ^2 are

$$\tilde{\mu} = \overline{X}$$
$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

To find the exact distributions of the method of moments estimators can prove to be difficult except if they are simple functions of statistics with known distributions as in the normal case where the method of moments estimators are simple functions of the MLE estimators [10].

5.1.2 Maximum likelihood estimation

The MLE estimators have more desirable properties than the method of moments estimators as exact distributions can often be found. If these distributions are difficult to derive we have asymptotic theory for univariate and multivariate parameters.

Let $(X_1, ..., X_n)$ be a random sample of random variable X which has a probability density function $f(x/\theta)$ for either the continuous or discrete case and parameter vector/scalar θ . By taking a random sample we ensure that the X_i are independent and identically distributed (i.i.d), the joint density of the observations, where for each random observation $X_i = x_i$ is observed, is then given by $f(x_1, ..., x_n/\theta)$ and is the product of the marginal densities $f(x_i/\theta)$.

$$\begin{aligned} f(x_{1,...,x_n}/\theta) &= \prod_{i=1}^{n} f(x_i/\theta) \\ &= L(\theta/x_1, ..., x_n) = L(\theta) \end{aligned} \tag{5.4}$$

Interpreting the joint density as a function of θ, given the realised observations $(x_1, ..., x_n)$, we call this function the likelihood function $L(\theta)$. We are interested in the theta that maximises the likelihood function, but rather than maximizing the likelihood function it is often easier to maximise the log of this function, which is the same thing since the logarithm is a monotone function. So for an i.i.d sample the log likelihood is.

$$\begin{aligned}\ell(\theta) &= \log L(\theta) = \log\left[\prod_{i=1}^{n} f(x_i/\theta)\right] \\ &= \sum_{i=1}^{n} \log[f(x_i/\theta)] \qquad (5.5)\end{aligned}$$

Example 5.2. *Normal Distribution:* Let $(X_1, ..., X_n)$ be a random sample from the normal distribution, then the likelihood is:

$$\begin{aligned}L(\mu, \sigma^2) &= f(x_1, ..., x_n/\mu, \sigma^2) \\ &= \prod_{i=1}^{n} f(x_i/\mu, \sigma^2) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/2\sigma^2} \qquad (5.6)\end{aligned}$$

The log likelihood is given by

$$\ell(\mu, \sigma^2) = -\frac{1}{2} n \log \sigma^2 - \frac{n}{2} \log 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \qquad (5.7)$$

Taking the partial derivatives of the log likelihood function with respect to μ and σ and setting them equal to zero, we get.

$$\begin{aligned}\frac{\partial \ell}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0 \\ \frac{\partial \ell}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} (x_i - \mu)^2 = 0\end{aligned}$$

Solving for μ and σ^2, the MLE's are

$$\begin{aligned}\hat{\mu} &= \overline{x} \qquad (5.8) \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2 \qquad (5.9)\end{aligned}$$

We have to remember that the estimator $\hat{\theta}$ that maximises the function $L(\theta)$ is a function of the realised observations, so $\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$, meaning the estimator is actually a random variable depending on the random variables

$(X_1, ..., X_n)$ rather than of their fixed numerical values x_i.

An estimator $\hat{\theta}$ of θ is called unbiased if $E(\hat{\theta}) = \theta$ and biased if $E(\hat{\theta}) \neq \theta$. Generally unbiased estimators are preferred to biased estimators, but biased estimators with smaller variances than unbiased estimators with large variances are sometimes preferred. The maximum likelihood estimator may be biased at times except in the case of very large n, as an example let us consider the estimate for σ^2 of the normal distribution compared to the actual equation.

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \neq \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \sigma^2 \tag{5.10}$$

Asymptotic properties of Maximum Likelihood Estimates

When the sample size becomes significantly large and under certain smoothness conditions of $f(x/\theta)$, such as the existence of the first two derivatives in the interval containing θ, maximum likelihood estimates have very desirable properties, we will state the theorems in this section and provide proves in the Appendix.

Theorem 3 (Theorem 5.1.) *Under smooth conditions of $f(x/\theta)$ the MLE $\hat{\theta}$ is consistent*[1]

Theorem 4 (Theorem 5.2.) [2] *If $f(x/\theta)$ is smooth then the MLE $\hat{\theta}$ converges to the normal distribution as $n \to \infty$.*

$$\hat{\theta} \xrightarrow{d} N(\theta_0, \frac{1}{nI(\theta_0)}) \quad as \quad n \to \infty$$

where θ_0 is the true value of parameter θ and $I(\theta)$ the asymptotic variance

[1] An estimate of θ based on a sample size of n is said to be consistent if the estimate $\hat{\sigma}_n$ converges in probability to θ as we increase n.

$$P(|\hat{\theta}_n - \theta| > \epsilon) \to 0 \quad as \quad n \to \infty$$

for a small finite ϵ

[2] see proof of theorem 5.2. in the Appendix

Confidence Intervals for Maximum Likelihood Estimates

From **Theorem 5.2.** we know that.

$$\hat{\theta} \xrightarrow{d} N(\theta_0, \frac{1}{nI(\theta_0)}) \quad as \quad n \to \infty$$
$$or \quad \sqrt{nI(\theta_0)}(\hat{\theta} - \theta_0) \sim N(0,1)$$

We can therefore give an asymptotic confidence interval for the true parameter θ_0 by

$$\theta_0 \in \hat{\theta} \pm z_{\alpha/2} \frac{1}{\sqrt{nI(\theta_0)}} \tag{5.11}$$

But since $I(\theta_0)$ in the above equation depends on the unknown parameter θ_0, we replace the value of θ_0 with the MLE estimate of θ namely $\hat{\theta}$ which is asymptotically an unbiased estimator of the parameter, and our asymptotic confidence interval for the parameter θ becomes.

$$\theta \in \hat{\theta} \pm z_{\alpha/2} \frac{1}{\sqrt{nI(\hat{\theta})}} \tag{5.12}$$

5.2 Generating Random Vectors

5.2.1 Conditional Multivariate Normal Distribution

Traditional risk management utilises the conditional normal multivariate distribution to generate Monte Carlo scenarios for the asset log-returns.

Let $\mathbf{X} = (x_1, ..., x_n)$ be the vector of n log-returns, where $x_{i,t-j+1} = lnP_{i,t-j+1} - lnP_{i,t-j}, j = 1, ..., T$ are T historical log returns and $P_{i,t-j}$ is the price of risk factor i at time $t-j$. The risk factors during the time step [t,t+1] are then sampled from the multinormal distribution $N(\mu_{t+1}, \Sigma_{t+1})$. The usual assumption[3] is that the mean return vector $\mu_{t+1} = 0$ and the elements of the $n \times n$ covariance matrix Σ_{t+1} are forecasted using the Exponentially Weighted Moving Averages (EWMA) method.

[3]e.g. Riskmetrics Technical Document of J.P.Morgan (1996)

EWMA

The standard formula for the daily variance estimates are $\Sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$, since we have assumed the population mean to be zero, we gain a degree of freedom and $\Sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$, letting $\alpha_i = \frac{1}{n}\forall i$ our variance estimate becomes $\Sigma_t^2 = \sum_{i<t}\alpha_i x_i^2$ where $\sum_i \alpha_i = 1$.

We would like the importance of an observation at time t to be λ times the observation at time $t+1$. λ is called the weight and we typically have $0.9 \leq \lambda \leq 1$. If we put $\alpha_{t-i} = \lambda^i$ and make the sum infinite we would have $\sum_{i=0}^{\infty}\alpha_{t-i} = \sum_{i=0}^{\infty}\lambda^i = \frac{1}{1-\lambda}$, so instead let $\alpha_{t-i} = (1-\lambda)\lambda^i$ and we get.

$$\sigma_{i,t+1}^2 = (1-\lambda)\sum_{j=0}^{N}\lambda^j x_{i,t-j}^2, \qquad i = 1,...,n \tag{5.13}$$

Where N are the historical observations used in the estimation procedure[4]. The covariance between risk factors i and j is:

$$\sigma_{i,j,t+1} = (1-\lambda)\sum_{k=0}^{N}\lambda^k x_{i,t-k}x_{j,t-k}, \qquad i,j = 1,...,n \tag{5.14}$$

Depending on the decay factor, more recent data have a higher influence in determining the estimated variance and covariances for the different risk factors.

If the matrix Σ is positive definite then there exists a matrix $\mathbf{B}$ s.t. $\Sigma = \mathbf{B}\mathbf{B}^T$ (Cholesky decomposition), if we now have a vector $\mathbf{Z} = (Z_1,...,Z_n)^T$ with each Z_i an independent standard normal variable, then the random vector $\mathbf{A} = \mu + \mathbf{B}\mathbf{Z}$ ($\mu \in \mathbb{R}^n$) is multivariate Gaussian with mean μ and covariance matrix Σ. We can use this procedure to forecast the log-returns for the conditional normal distribution.

5.2.2 Copula

The main aim of the Copula function is to be able to generate dependent uniform variates. For a given Copula $\mathbf{C}$ we can achieve this with the following general

[4] Riskmetrics and [16] use N=74 and λ is assumed to be 0.94

algorithm for N dependent variables:

1. Generate N independent *uniform* variates $(v_1, ..., v_N)$
2. Recursively generate the N variates in the following manner

$$u_n = C^{-1}_{u_1,\dots,u_n}(v_n) \tag{5.15}$$

where

$$\begin{aligned} C_{u_1,\dots,u_n}(u_n) &= \mathbb{P}[U_n \le u_n | (U_1, ..., U_{n-1}) = (u_1, ..., u_n)] \\ &= \frac{\partial^{n-1}_{(u_1,\dots,u_{n-1})} C(u_1, ..., u_n, 1, ..., 1)}{\partial^{n-1}_{(u_1,\dots,u_{n-1})} C(u_1, ..., u_{n-1}, 1, ..., 1)} \end{aligned} \tag{5.16}$$

In this manner we generate each u_n using its conditional distribution.

Empirical copulas

Definition 4: *The empirical distribution function for a random sample $X_1, ..., X_n$ from F_X is the function*

$$\hat{F}_{X_1,\dots,X_n}(t) := \frac{1}{n}\#[i, 1 \le i \le n : X_i \le t]$$

The strong law of large numbers (S.L.L.N.) gives point wise convergence if the random sample is drawn such that the X_i are independent:

$$\forall t \in \mathbb{R} : \lim_{k\to\infty} \hat{F}_{X_1,\dots,X_n}(t) = F_X(t) a.s.$$

Now let $\Im = (x_1^t, ..., x_N^t)_{t=1}^T$ denote a sample of N observations at a time t, then the empirical copula distribution is given by

$$\hat{C}\left(\frac{t_1}{T}, ..., \frac{t_N}{T}\right) = \frac{1}{T}\sum_{i=1}^{T} \mathbf{1}_{[x_1^t \le x_1^{(t_1)}, \dots, x_N^t \le x_N^{(t_N)}]} \tag{5.17}$$

where the $x_n^{(t)}$ are the order statistics of variable X_n and $1 \le t_1, ..., t_N \le T$. The empirical copula frequency can be defined as

$$\hat{C}_f(\frac{t_1}{T},...,\frac{t_N}{T}) = \frac{1}{T} \ if \ (x_1^{(t_1)},...,x_N^{(t_N)}) \in \Im \ or \ 0 \ otherwise. \tag{5.18}$$

We then get the following relationships between the empirical copula distribution and the empirical copula frequency [13]:

$$\hat{C}(\frac{t_1}{T},...,\frac{t_N}{T}) = \sum_{i_1=1}^{t_1} \cdots \sum_{i_N=1}^{t_N} \hat{C}_f(\frac{i_1}{T},...,\frac{i_N}{T}) \tag{5.19}$$

and

$$\hat{C}_f(\frac{t_1}{T},...,\frac{t_N}{T}) = \sum_{i_1=1}^{2} \cdots \sum_{i_N=1}^{2} (-1)^{i_1+...+i_N} \hat{C}(\frac{t_1 - i_1 + 1}{T},...,\frac{t_N - i_N + 1}{T}) \tag{5.20}$$

When choosing a copula the copula which minimises the difference, based on the L^n norm, between the empirical copula distribution and the chosen copula should be used, so let $C_{k1\leq k\leq p}$ be a set of available p copula then we find:

$$\overline{d}_n(\hat{C}, C_k) = \left(\sum_{t_1=1}^{T} \cdots \sum_{t_n=1}^{T} \left[\hat{C}\left(\frac{t_1}{T},...,\frac{t_n}{T}\right) - C_k\left(\frac{t_1}{T},...,\frac{t_n}{T}\right)\right]^2 \right)^{\frac{1}{2}} \tag{5.21}$$

The distance can also be used to estimate the parameters for a copula C_k. For a vector of parameters $\chi \in \aleph$ we have:

$$\hat{\chi} = arg \min_{\chi \in \aleph} \left(\sum_{\mathbf{u} \in \ell} \left[\hat{C}(\mathbf{u}) - C(\mathbf{u}; \chi) \right]^2 \right)^{\frac{1}{2}} \tag{5.22}$$

Gaussian Copula

The Gaussian copula is the copula of a multivariate normal distribution and a vector $\mathbf{X} = (X_1, ..., X_n)$ is multivariate normal iff:

1. The univariate margins $F_1, ..., F_n$ are Gaussian and
2. The dependence structure among the marginals is described by the unique copula function:

$$C_{\mathbf{R}}^{Ga}(u_1,...,u_n; \mathbf{R}) = \Phi_{\mathbf{R}}(\varphi^{-1}(u_1),...,\varphi^{-1}(u_n)) \tag{5.23}$$

Where $\Phi_{\mathbf{R}}$ is the standard multivariate normal c.d.f. with a linear correlation matrix $\mathbf{R}$ and ϕ^{-1} is the inverse standard univariate Gaussian distribution.

Before we discuss a method for determining the parameter $\mathbf{R}$ for the Gaussian copula. the historical data have to be independent determinations from a common c.d.f. So we filter the data using variances calculated with the EWMA method.

$$z_{i,t-j} = \frac{x_{i,t-j}}{\sigma_{i,t-j}}, \qquad i = 1, ...,; \quad j = 0, ..., T - 74 \tag{5.24}$$

Now to estimate the parameter $\mathbf{R}$ for the Gaussian copula, [16]:

1. transform the dataset of standardised log-returns (5.24) $(z_1^t, ..., z_n^t), t = 1, ..., T-74$ into uniform variates $(\hat{u}_1^t, ..., \hat{u}_n^t)$ using the marginal empirical distributions or other selected marginal distributions.

2. Using the inverse standard Gaussian distribution Φ transform the historical data further to : $\varsigma_t = (\Phi^{-1}(\hat{u}_1^t), ..., \Phi^{-1}(\hat{u}_n^t)), t = 1, ..., T-74$

3. Finally we calculate $\mathbf{R}$ using this transformed dataset ς_t and the EWMA method with equations (5.13),(5.14)

To generate random variates from the Gaussian copula we will use the following procedure:

1. Find the Cholesky decomposition B of R.

2. Generate n independent standard normal random variates $\mathbf{Z} = (z_1, ..., z_n)^\top$.

3. Then let $\mathbf{X} = \mathbf{BZ}$, the x_i are then correlated standard normal data.

4. Transform the x_i using the standard normal distribution function ϕ, $u_i = \phi(x_i), i = 1, ..., n$.

5. The final vector $(u_1, ..., u_n)^\top$ contains the random variates from the n dimensional Gaussian copula $C_{\mathbf{R}}^{Ga}$.

Student's t-copula

For the Student's t-copula it is not possible to obtain an analytical expression for the maximum likelihood estimate parameter $\hat{R}_{ML}$, but we can construct an algorithm that we repeat until we get convergence.

The likelihood function for the student t-copula is given by:

$$\begin{aligned} \ell(\theta) &= T(\ln\Gamma(\frac{\nu+N}{2}) - \Gamma(\frac{\nu}{2})) - NT(\Gamma(\frac{\nu+1}{2}) - \Gamma(\frac{\nu}{2})) - \frac{T}{2}\ln|R| \\ &- (\frac{\nu+N}{2})\sum_{t=1}^{T}\ln(1+\frac{1}{\nu}\varsigma_t R^{-1}\varsigma_t^{\top}) + (\frac{\nu+1}{2})\sum_{t=1}^{T}\sum_{n=1}^{N}\ln(1+\frac{\varsigma_n^2}{\nu}) \end{aligned}$$

Where T is the number of observations, N the number of instruments and ν the degrees of freedom. Maximizing the above function w.r.t. R^{-1} we get:

$$\frac{\partial\ell(\theta)}{\partial R^{-1}} = -\frac{T}{2}R - (\frac{\nu+N}{2})\sum_{t=1}^{T}\frac{\frac{1}{\nu}\varsigma_t^{\top}\varsigma_t}{1+\frac{1}{\nu}\varsigma_t R^{-1}\varsigma_t^{\top}} \tag{5.25}$$

So the ML estimate R_{ML} must satisfy the following non-linear matrix equation:

$$\hat{R}_{ML} = \frac{1}{T}(\frac{\nu+N}{\nu})\sum_{t=1}^{T}\frac{\varsigma_t^{\top}\varsigma_t}{1+\frac{1}{\nu}\varsigma_t\hat{R}_{ML}^{-1}\varsigma_t^{\top}} \tag{5.26}$$

We obtain the above mentioned $\hat{R}_{ML}$ using the following algorithm:

1. let $\hat{R}_0$ be the $\hat{R}_{ML}$ matrix obtained for the gaussian copula.
2. let $\hat{R}_{m+1}$ be obtained in the following way

$$\hat{R}_{m+1} = \frac{1}{T}(\frac{\nu+N}{\nu})\sum_{t=1}^{T}\frac{\varsigma_t^{\top}\varsigma_t}{1+\frac{1}{\nu}\varsigma_t\hat{R}_m^{-1}\varsigma_t^{\top}} \tag{5.27}$$

3. Repeat step 2 until we obtain convergence.
4. $\hat{R}_{m+1} = \hat{R_m}(:= \hat{R}_{\infty})$

To generate random variates from the Students t-copula we will use the following procedure:

1. Find the Cholesky decomposition B of R.
2. Generate n independent standard normal random variates $\mathbf{Z} = (z_1, ..., z_n)^\top$.
3. generate a random variate, s, from a χ^2_ν distribution independent of $\mathbf{Z}$
4. Then let $\mathbf{Y} = \mathbf{BZ}$.
5. Next we set $\mathbf{X} = \frac{\sqrt{\nu}}{\sqrt{s}}\mathbf{Y}$.
6. Transform the x_i using the univariate students t-distribution function t_ν, $u_i = t_\nu(x_i), i = 1, ..., n$.
7. The final vector $(u_1, ..., u_n)^\top$ contains the random variates from the n dimensional Students t-copula $C^t_{\nu,\mathbf{R}}$.

In order to obtain K scenario values for the log-returns $\mathbf{X}$, we transform the variates obtained above using the respective marginal distribution for each risk factor, $\mathbf{Z} = (z_1, ..., z_n)^\top = \left(F_1^{-1}(u_1), ..., F_1^{-1}(u_1)\right)$, and then rescale these standardised log-returns using the EWMA variances, $\mathbf{X} = (x_1, ..., x_n)^\top$ $= (z_1\sigma_{1,t+1}, ..., z_n\sigma_{n,t+1})^\top$

Using these K scenarios we re-evaluate the portfolio for n instruments at time t+1 as:

$$P_{j,t+1} = \sum_{i=1}^{n} P_{i,t}exp(x_{i,j}), \quad j = 1, ..., K \tag{5.28}$$

And the portfolio losses[5] for each scenario:

$$\begin{aligned} L_j &= P_t - P_{j,t+1} = \sum_{i=1}^{n}[P_{i,t} - P_{i,t}exp(x_{i,j})] \\ &= \sum_{i=1}^{n} P_{i,t}(1 - exp(x_{i,j})), \quad j = 1, ..., K \end{aligned} \tag{5.29}$$

Where $P_{i,t}$ is the market price of instrument i at time t.

[5] profits are negative losses

Chapter 6

Case Study: Johannesburg Stock Exchange

The data is collected from the *Troskie* database of Johannesburg Stock Exchange (JSE) stock prices, from the Department of Statistical Sciences at the University of Cape Town, and ranges from 1986-12-06 to 2003-01-25 (843 prices). Monthly and weekly stock prices are available, to increase the number of data points we perform our analysis on the weekly stock prices.

We select six stocks for which we could find extensive data to set up our CVaR model and to perform the backtesting of the model. The six stocks are: ABSA, Anglo Gold, Goldfields, Harmony, Bidvest and BarWorld.

6.1 CVaR Model

Necessary requirements for the model are that it should be easy to set up, use and modify. Visual basic for applications is used in the Excel environment since it can be accessed in almost any office environment. Using the historical time series of stock prices the CVaR algorithm is as follows:

1. Calculate the time series of log-returns (842 data points);
2. Calculate the EWMA variance and covariance with $\lambda = 0.94$ and $N = 74$;
3. Generate n random variates using the t-copula following the algorithm in section (5.2.2);

4. Repeat step 3 to generate j scenarios;
5. Using these scenarios, re-evaluate the portfolio and the portfolio losses using equation (5.28) and (5.29);
6. Specify the quantile for which the CVaR is to be calculated, ζ ;
7. optimise the portfolio and calculate the ζCVaR and ζVaR of the portfolio using algorithm (3.17).

The last step of the algorithm is executed using Microsoft Office Excel solver[1] and we set the solver options as follows[2]:

- Max time = 1000 seconds;
- Iterations = 30000;
- Precision = 0.000001;
- Tolerance = 5%;
- Convergence = 0.000001;
- Assume non-negative model: Causes Solver to assume a lower limit of zero for all adjustable cells for which the user has not set a lower limit (this is the only model that could start the solver for all scenario sizes) for this model we specify
  - Estimates = Tangent;
  - Derivatives = Central;
  - Search = Newton.

6.2 Setting up the Experiments

6.2.1 Determining the number of scenarios for a stable CVaR solution

The set of scenarios generated has to be large enough to give a consistent and stable CVaR solution, but should not be computationally expensive and time

[1]The Microsoft Excel Solver tool uses the Generalized Reduced Gradient (GRG2) nonlinear optimization code developed by Leon Lasdon, University of Texas at Austin, and Allan Waren, Cleveland State University.- From Excel Help

[2]We found through trial and error that these settings were the best to cope with optimising a large number of scenarios

consuming. In this section we will find the number of scenarios necessary to satisfy these criteria.

We use our CVaR model from the previous section and for simplicity assume that the log-returns of the instruments have normal marginal distributions as well as a t-Copula with $v = 3$. The CVaR model is repeated while the number of scenarios are increased each time starting from 50 until the CVaR minimization value does not change. The number of scenarios at this point will be the minimum that can be used for a consistent CVaR minimization model. The one week scenarios are generated for t plus the last data point 2003-01-25.

The maximum amount of the portfolio one of the six instruments is allowed to be is 30% and the minimum is 0%, we therefore add an additional constraint to our CVaR algorithm, $0 \leq w_i \leq 0.30$.

We also time the scenario generation to get a sense of the computational demands of the models scenario generation algorithm using a computer with a 1.60 GHZ processor and 256 MB of RAM.

6.2.2 Selecting the Marginal t-Distributions and t-Copula

The question of which marginal and t-Copula (v parameter) distribution best fit our model of current stocks, will be answered using a bottom-up approach. This means that we will first determine the t-distributions that best fit the empirical distributions of each stock, using these marginal distributions we will then determine the t-Copula which best describes the dependence structure between the instruments in the portfolio.

In order to ensure that we have chosen the best fit t-distributions for the n instruments we use QQ-plots[3] to compare the distribution of loss scenarios obtained using various t-distributions to that of the actual historical losses.

In figure(6.1) the QQ-plot on the left is a plot of the same stock (ABSA)

[3] QQ-plots display a quantile-quantile plot of two samples. If the samples are from similar distributions, the plot will be linear

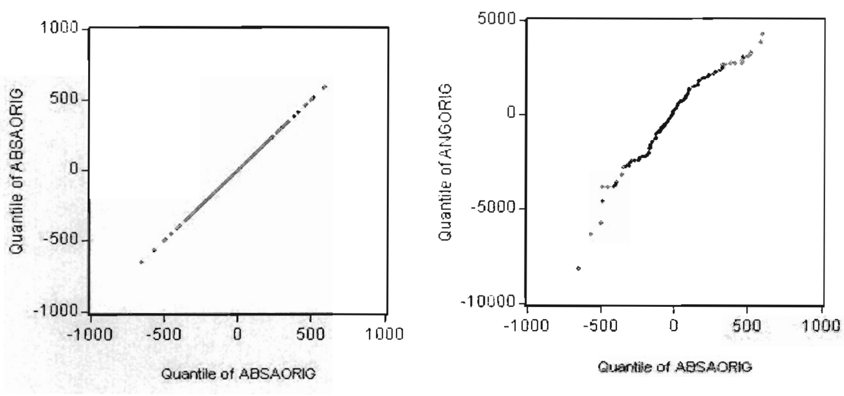

Figure 6.1: QQ-Plot Example

and is therefore linear, while the QQ-plot on the right is a plot of two different stocks (ABSA versus Anglo) and is therefore not a perfect linear plot. We want to find the t-distribution and t-Copula which gives the best linear QQ-plot.

For the marginal distributions, the first four steps of the CVaR model are used, only for $n = 1$ instrument, to generate 500 scenarios that are used for the QQ-plot. Of the 500 scenarios that are generated 150 are randomly selected and a QQ-plot is plotted against the most recent 150 gains/losses of the stock even though the EWMA covariance matrix that is calculated in the CVaR model only makes use of the most recent 74 data points.

We do this to verify how well we are capturing historical returns that fall away during the EWMA covariance matrix calculation for the scenario generation process. This is repeated for a number of different t-distributions with different degrees of freedom, v.

After selecting the marginal distributions with the best linear QQ-plots, we repeat a similar experiment for the best t-Copula. The first four steps of the CVaR model are used to generate 500 scenarios, for the $n = 6$ instruments, that are used for the QQ-plot. Of the 500 scenarios that are generated 150 are randomly selected and a QQ-plot is plotted against the most recent 150 gains/losses for each stock. This is repeated for a number of different t-Copula with different degrees of freedom, v.

The one week scenarios are generated for t plus the last data point 2003-01-25.

6.2.3 Back-testing the CVaR Model

To test our model we must back-test the model, this means we calculate the $\zeta - CVaR$ of our portfolio of n instruments for k one week time periods. If the percentage of actual losses that are greater than the optimised $\zeta - CVaR$ threshold during the k periods is greater than $1 - \zeta$, then our CVaR model is not a good indicator of the maximum loss our portfolio can suffer over a one week time horizon for a given confidence level.

For the back-testing of our model we back-test 100 successive weeks starting at 2001-03-03 and ending at 2003-01-25, the number of scenarios generated for the back-testing model will depend on the minimum number of scenarios for a stable CVaR solution due to the time consuming nature of the back-testing procedure.

We back-test for $\zeta = 0.95$ and $\zeta = 0.99$ calculating CVaR and 3 VaR indicators; α from the CVaR minimization, Parametric VaR[4] (equation(3.2)) and Non-Parametric/Monte-Carlo VaR[5].

6.2.4 Efficient Frontier

Using the CVaR model we will generate an efficient frontier for our portfolio based on the same premise as that of modern portfolio theory, meaning it will be based on a risk/return profile. The only difference will be our definition of risk, which will be the maximum amount[6] of the original equally weighted portfolio the investor is willing to lose, R_{CVaR}. Our model must therefore optimise the portfolio so that the $\zeta - CVaR$ is always less than or equal to R_{CVaR}.

[4] The covariance matrix is calculated using EWMA variances

[5] The portfolio (X-vector) for the scenario generation of losses is similar to that obtained from the CVaR minimisation for consistency

[6] This amount is specified by the user as a percentage of the original portfolio

For completeness we will generate efficient frontiers using modern portfolio theory (CAPM and Markowitz models), the efficient frontiers are generated for a one week period ending 2003-01-25, t plus the second last data point 2003-01-18.

Defining the loss over the one week period as the average of j loss scenarios, the CVaR minimization algorithm for the efficient frontier then becomes:

$$\begin{aligned} \min_{x,\alpha} & \quad \frac{\sum_{i=1}^{j} f(x, y_i)}{j}. \\ \hat{F}_{\varsigma}(\mathbf{x}, \alpha) & \leq R_{CVaR} \\ subject & \quad to \\ \sum_{i=1}^{n} w_i & = 1, 0 \leq w_i \leq 0.30, i = 1, ..., n \end{aligned} \tag{6.1}$$

We repeat the minimization algorithm for increasing CVaR risk levels, $R_{CVaR} = \{3\%, ..., R_n\}$ until the maximised return from the CVaR model levels off. The original portfolio that the CVaR risk limits are applied to is an equally weighted portfolio for the 6 stocks in the portfolio. The efficient frontier is calculated for 0.95-CVaR and 0.99-CVaR.

The Markowitz and CAPM models are also used to generate efficient frontiers for our portfolio of 6 stocks. We specify a target return and minimise the resulting portfolio variance. The frontier is created for the date of 2003-01-25, the data set consists of $k = 144$ weekly log-returns of the 6 stocks (2000-04-22 to 2003-01-25).

Using the methodology of section 2.1, we use the selected dataset to calculate the mean returns of the stocks which are used as the expected return of the stocks and also to calculate the covariance matrix used to determine the portfolio variance for the Markowitz model. We do not use the EWMA variance, instead we use the traditional formula $covar = \frac{1}{k}\sum_{i=1}^{k}(x_i - \bar{x})(y_i - \bar{y})$ and $var = \frac{1}{k-1}\sum_{i=1}^{k}(x_i - \bar{x})^2$.

For the single model CAPM portfolio optimization, section 2.2, the JSE-Overall Index is selected as the proxy for the South African market. The data set, similar range to that of the 6 stocks, is used to calculate the alpha and beta coefficients for the model.

The Markowitz and CAPM frontiers will consist of the same number of points as the CVaR efficient frontiers.

6.3 Findings and Discussion

6.3.1 Determining the number of scenarios for a stable CVaR solution

	95%		99%	
Scenarios	CVaR	VaR	CVaR	VaR
50	532.29	464.18	626.88	626.88
250	510.38	378.86	669.43	610.10
500	518.98	369.64	674.46	626.89
1000	482.95	340.89	701.25	654.93
1500	481.06	338.96	688.22	603.86
2000	479.03	344.74	688.26	605.12
2500	483.57	344.93	686.13	599.52
3000	481.40	341.51	688.59	602.84

Table 6.1: Scenario Stability Table

The results for the scenario stability are summarised in table (6.1) and plotted in figure (6.2).

The top graph in figure (6.2) plots the value of α or VaR obtained from the CVaR minimization algorithm for increasing scenarios and 2 confidence levels. For $\zeta = 0.95$, as the number of scenarios increase the VaR value decreases and stabilises at around 340 cents above 1000 scenarios. The VaR value for $\zeta = 0.99$ increases, decreases and stabilises at around 600 cents above 1500 scenarios.

For CVaR stability we find that the CVaR value at $\zeta = 0.95$ decreases and stabilises above 1000 scenarios at around 482 cents. The $\zeta = 0.99$ CVaR value increases as the number of scenarios increase and stabilises above the 1500 scenario mark at a value of around 688 cents.

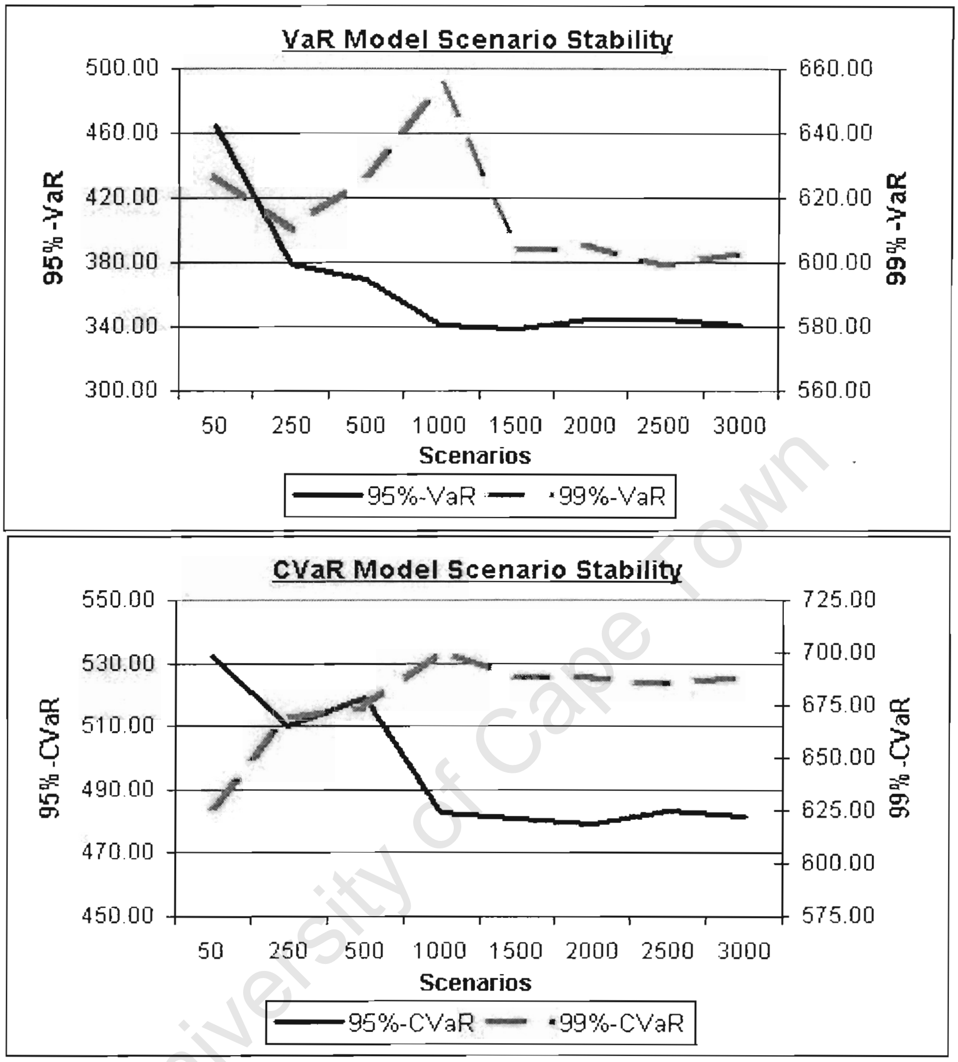

Figure 6.2: Scenario Stability Graphs

The values for VaR and CVaR both stabilise at a lower number of scenarios for the lower confidence level of $\zeta = 0.95$, the model therefore requires a higher number of scenarios to minimise the CVaR algorithm if the user increases the confidence level for which the portfolio optimization is being run. In this case the minimum number of scenarios we can use for model stability at both the $\zeta = 0.95$ and $\zeta = 0.99$ confidence levels are 1500.

Scenarios	Seconds	Minutes
500	140.74	2.35
1000	170.50	2.84
2000	235.54	3.93
4000	348.37	5.81
6000	730.52	12.18
8000	1084.78	18.08
10000	1459.97	24.33

Table 6.2: Scenario Time Table

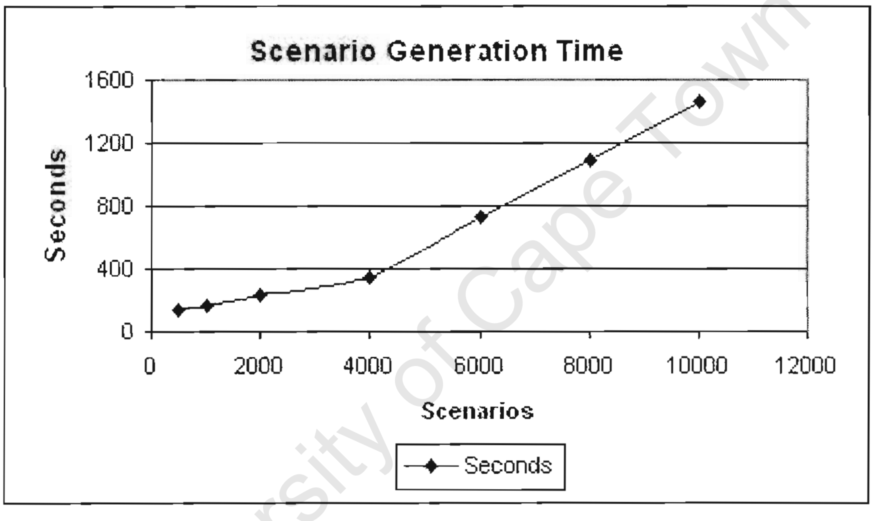

Figure 6.3: Scenario Time Graph

The time taken to generate scenarios increases more rapidly after 4000 scenarios and has a linear trend before and after this transition point.

From the graph we can calculate the time taken to complete the scenario generation to backtest our model and to optimise the time, in terms of scenario generation, for other investigations.

6.3.2 Selecting the Marginal t-Distributions and t-Copula

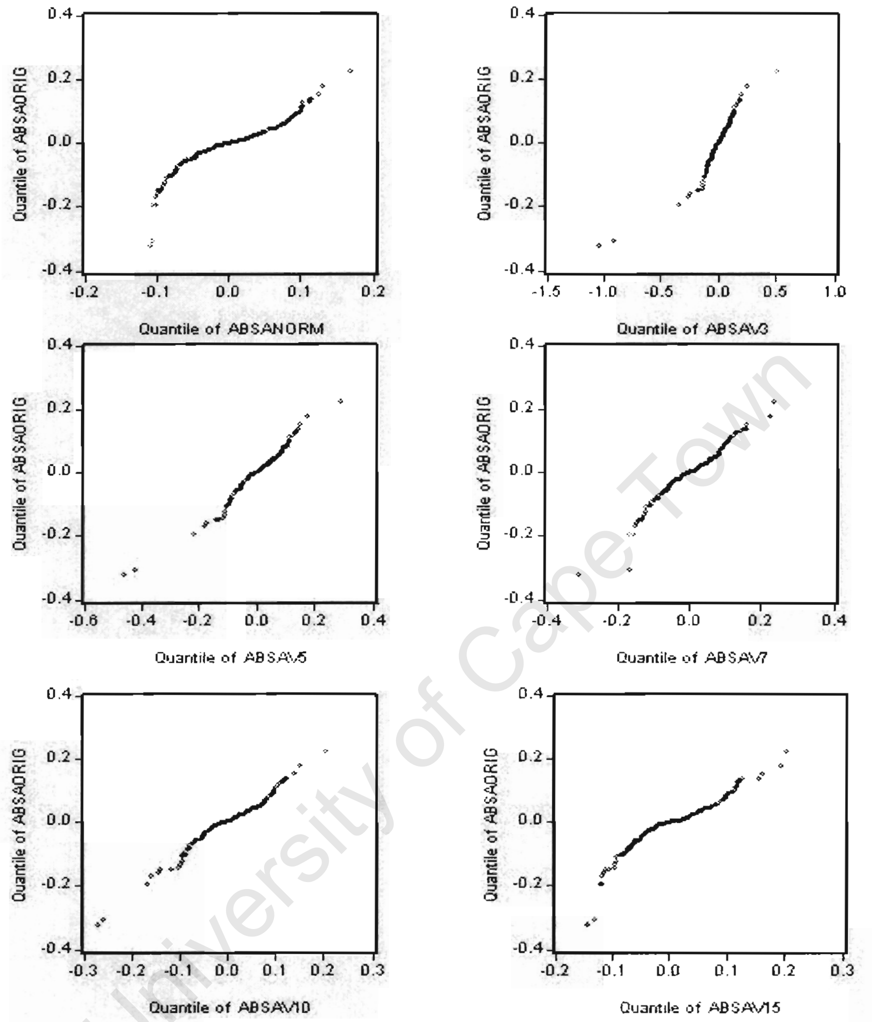

Figure 6.4: Absa T-Marginals

For each stock five different t-marginal distributions are considered, $t = \{3, 5, 7, 10, 15\}$. we also use a standard normal distribution.

The QQ-plots in figure (6.4) are for scenarios from the different marginal distributions of the *Absa* stock. We see that the QQ-plots from the log-return scenarios generated using $t = 5, 7$ and 10 marginal distributions generate the best linear fit to the actual *Absa* log-returns.

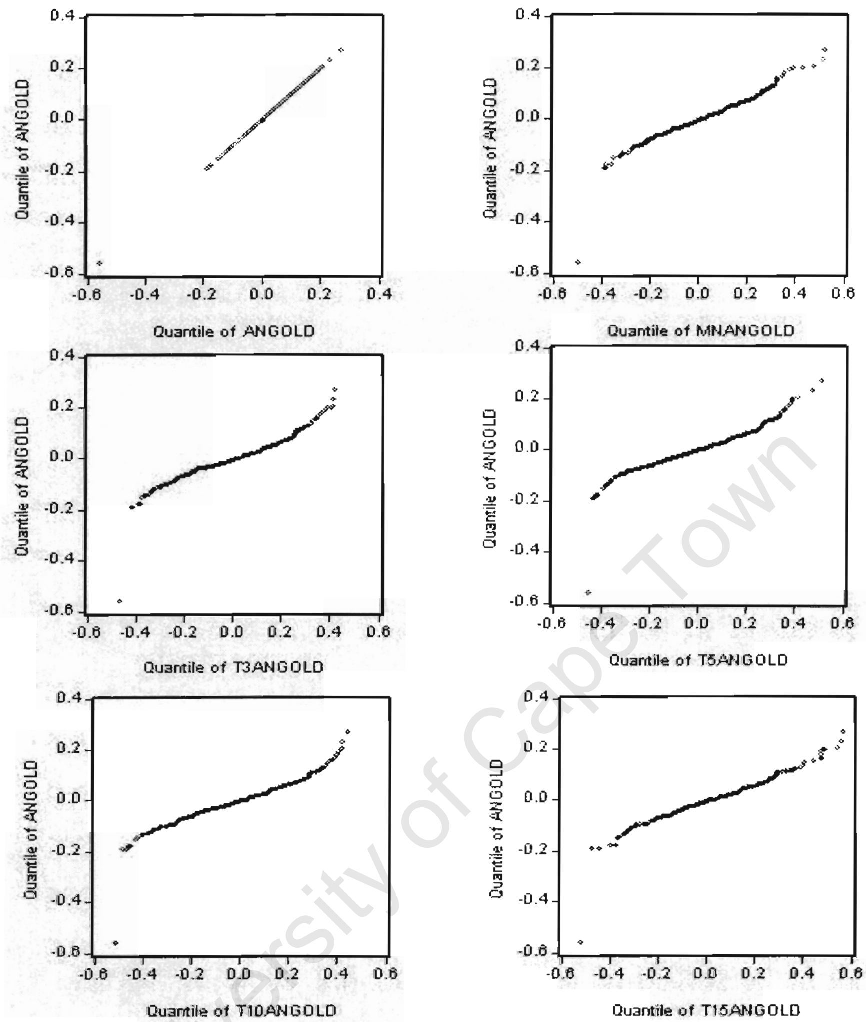

Figure 6.5: Anglo T-Copula

Studying the QQ-plots for the other 5 stocks, we found that the t=7 marginal distribution is common among all the stocks as one of the distributions that generates the best linear fit to the historical log-returns. we will therefore use this marginal distribution as the best fit for the six stocks in our model.

After selecting the marginal distributions, we repeat similar plots for the copula which best describes the dependence characteristics among the six stocks, we consider the normal copula plus 4 t-copula, $t = \{3, 5, 10, 15\}$.

Figure (6.5) shows the QQ-plots for the *Ang* stock, we find that the $t = 3$ copula best describes the dependence structure between the 6 stocks.

Our model will therefore use $t = 7$ marginal distributions for the 6 stocks and a $t = 3$ copula.

6.3.3 Back-testing the CVaR Model

ζ=0.95	
CVaR	VaR/alpha
5.00%	15.00%
Parametric VaR	Non-Parametric VaR
8.00%	15.00%

ζ=0.99	
CVaR	VaR/alpha
1.00%	4.00%
Parametric VaR	Non-Parametric VaR
4.00%	4.00%

Table 6.3: Actual Losses Exceeding Monetary Risks

Taking the model scenario stability results of section (6.3.1) into consideration, we decided to used 2000 scenarios for the back-testing of our model, in an effort to save computational time. Table (6.3) shows the number of actual portfolio losses in the 100 week period that exceed the thresholds we have set using our CVaR model.

For the 95% confidence level, the CVaR value is only exceeded 5% of the time, while the VaR/alpha from the CVaR model and the Non-parametric/Monte-Carlo VaR are Exceeded 14% of the time and the parametric VaR is exceeded 8% of the time. The 99% confidence level results show that the CVaR value is exceeded 1% of the time while the other VaR indicators are all exceeded 4% of the time.

We found that the CVaR value is always greater than that of the 3 VaR indicators during our test period. The CVaR value will therefore always be one of the indicators which is exceeded the least by actual portfolio losses.

The graphs in figure (6.6) show the CVaR and VaR/alpha envelopes covering the actual gains/losses during the back-testing period. To better understand

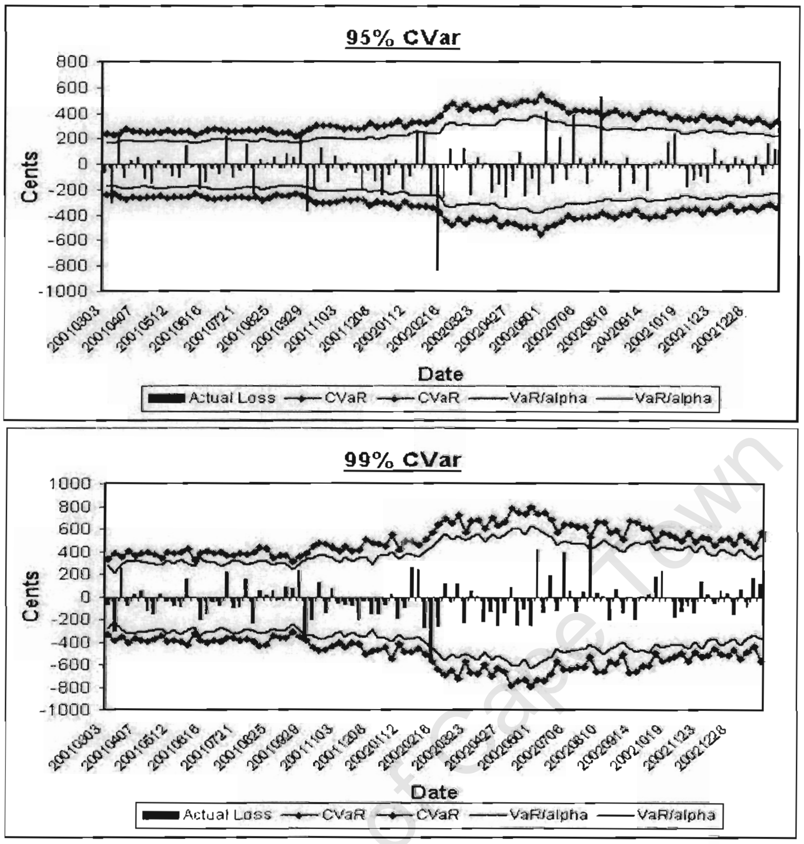

Figure 6.6: CVaR Model Losses

the increase in envelope size during the mid-section of the back-testing period we plot the EWMA variance of the 6 stocks during the same time period. We notice an increase in volatility for 4 of the 6 stocks at the same time the envelope size starts to increase.

We can therefore draw some comfort from the fact that the model captures increases in market volatility, to answer the question of whether or not it may be capturing the volatility too late we study the dates at which the actual losses exceed the CVaR threshold. For the 95% confidence level the first 2 dates are at the start of the backtesting period (second and third backtesting point), the third date is before the increase in market volatility and expansion of CVaR envelope, the fourth occurs as the increase in envelope sise and market volatility starts, while the final date occurs as the envelope is converging.

Studying the 99% confidence level graph we find that the date when the actual portfolio loss exceeds the CVaR threshold occurs after the envelope has developed and is busy converging[7].

We should also note that the largest actual portfolio returns/losses occur during the increase in market volatility which is also the period when the increase in envelope size occurs.

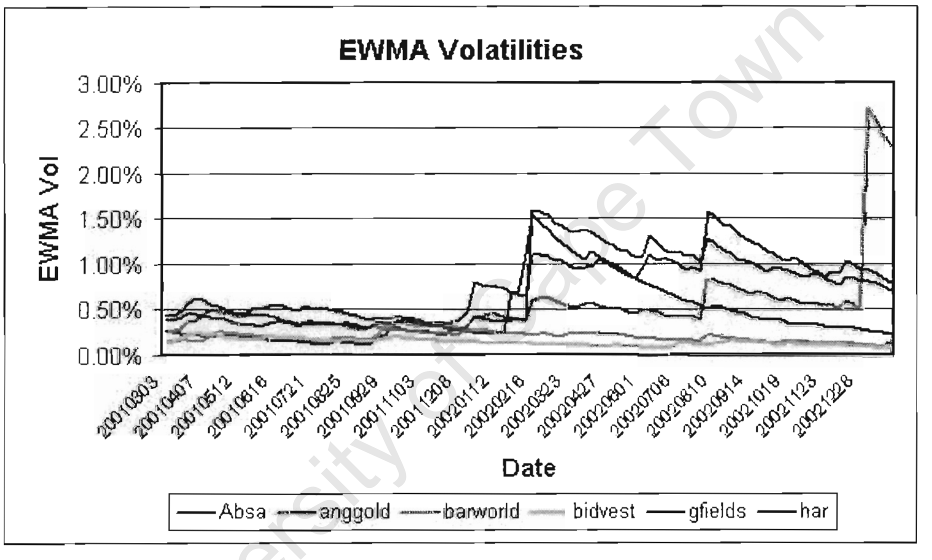

Figure 6.7: EWMA Volatilities

[7]note that as the EWMA variance of the JSE-Overall index is decreasing the envelope also decreases in sise

6.3.4 Efficient Frontier

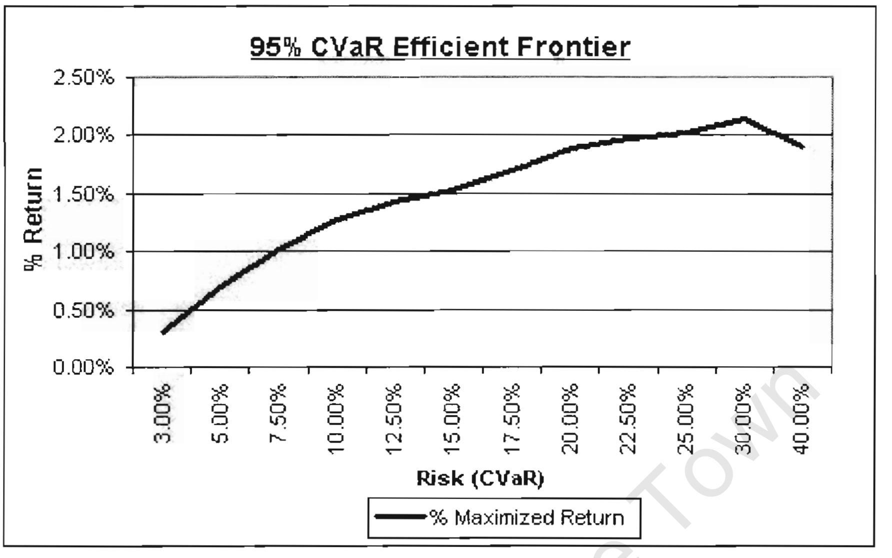

Figure 6.8: 95% Efficient Frontier

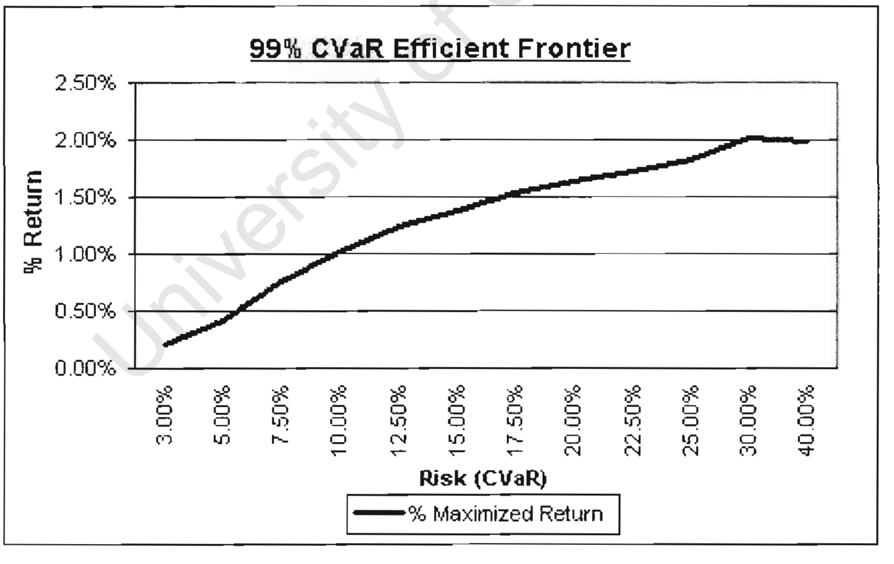

Figure 6.9: 99% Efficient Frontier

For the efficient frontier we generate 10000 scenarios for the CVaR minimization algorithm to ensure model stability, the level of risk, R_{CVaR}, is increased until the maximised return starts leveling off.

Studying the shape of the CVaR efficient frontiers, we see that both graphs develop in a concave manner. The maximised returns for both confidence levels increase as we increase the fraction of the original portfolio that we are willing to lose, this increase in maximised returns starts leveling off at the 30% R_{CVaR} level and a maximim maximised return of 2.00%, once again for both confidence levels.

We plot the maximised returns versus the maximum allowed risk R_{CVaR} in figures (6.8) and (6.9). The errors , which are the difference between the maximised return of the CVaR model and the actual portfolio one week return[8], are plotted in figure (6.10). The CVaR frontiers are made up of 12 risk-return points.

The maximised returns for both confidence levels increase as we increase the fraction of the original portfolio that we are willing to lose, this increase in maximised returns starts leveling off at the 30% R_{CVaR} level and a maximim maximised return of 2.00%, once again for both confidence levels.

It is interesting to note that the error changes from positive to negative between 30% and 40% just after the point when the maximised return starts leveling off at the 30% R_{CVaR} mark, this means that the investor would have made a higher profit than predicted by the model had they invested according to the proposed allocations.

If we look at the errors[9] of the CAPM and Markowitz frontiers we see that 92% and 83% of the errors, respectively, are negative. The investor would have therefore realised a profit higher than what the models predicted, and therefore most of the actual portfolio returns are not contained within the efficient frontiers of these two models. The percentage of negative errors for the 95% and 99% CVaR models are 8% each.

[8] Given that the portfolio weights at the beginning of the period are similar to used for the CVaR minimised portfolio

[9] See efficient frontier tables at the end of this section

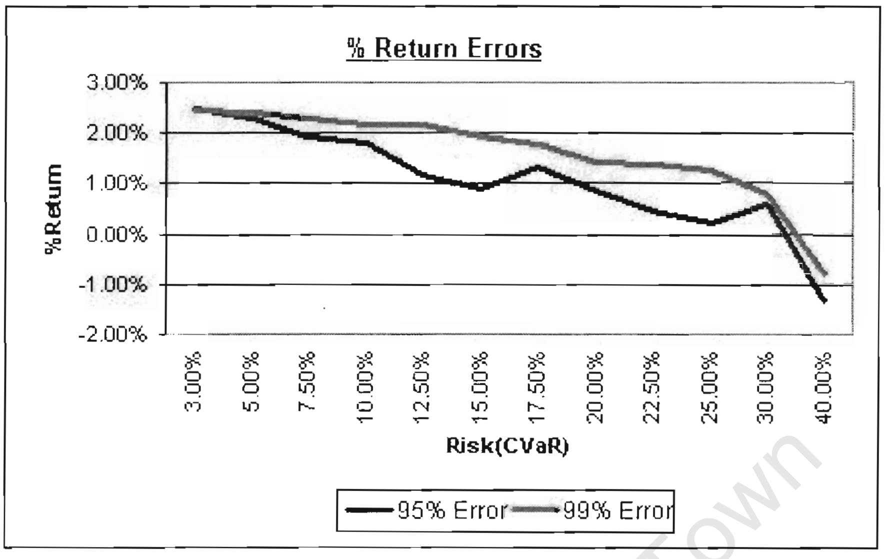

Figure 6.10: Frontier Errors

Before the reversal in the error value the 99% CVaR model shows a less volatile error than the error generated by the 95% CVaR model. meaning a lower tracking error[10]. This is due to the fact that the actual portfolio returns[11] for the 99% CVaR frontier increases as the R_{CVaR} value increases while the actual portfolio returns for the 95% CVaR frontier fluctuates up and down as we increase the R_{CVaR} value.

Tracking Errors			
99% CVaR	95% CVaR	Markowitz	CAPM
0.92%	1.04%	0.72%	0.62%

Table 6.4: Tracking Error

The CAPM frontier has the lowest tracking error, followed by the Markowitz frontier, the 99% CVaR frontier and finally the 95% CVaR frontier. This means

[10]We calculate tracking error by calculating the standard deviation of the error which is the difference between the maximised returns and actual portfolio returns.

[11]See efficient frontier tables at the end of this section

that the error generated by the CAPM model has been the most consistent and we can use the average of these errors to adjust the expected returns, from the optimisation procedure,to give the closest estimate of the actual portfolio returns for the given portfolio allocations.

The Markowitz and CAPM models both exhibit the characteristic concave shape.

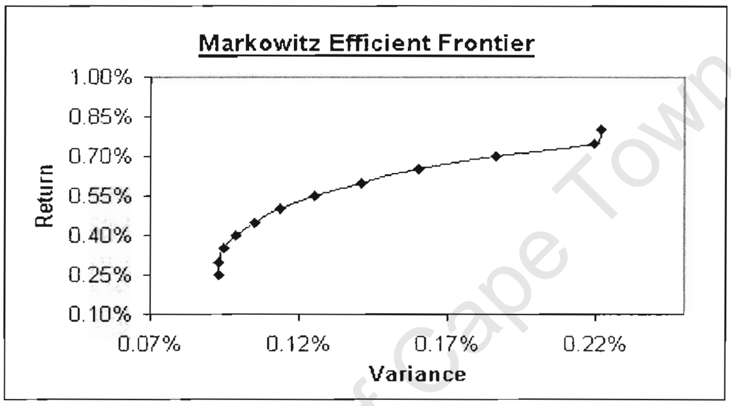

Figure 6.11: Markowitz Frontier

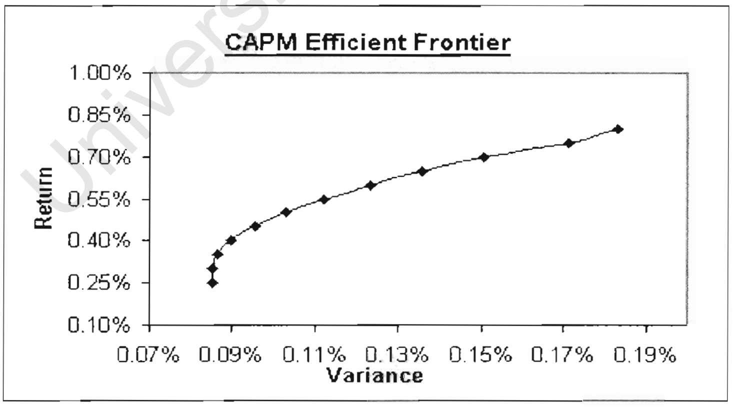

Figure 6.12: CAPM Frontier

Risk (w)	% Maximized Return	% Actual Portfolio Return	Error
3.00%	0.30%	-2.19%	2.49%
5.00%	0.70%	-1.60%	2.30%
7.50%	1.02%	-0.93%	1.94%
10.00%	1.28%	-0.53%	1.81%
12.50%	1.42%	0.26%	1.16%
15.00%	1.53%	0.62%	0.90%
17.50%	1.70%	0.37%	1.33%
20.00%	1.90%	1.03%	0.86%
22.50%	1.97%	1.50%	0.47%
25.00%	2.03%	1.79%	0.24%
30.00%	2.14%	1.55%	0.60%
40.00%	1.90%	3.22%	-1.33%

	95% CVaR Portfolio					
Risk (w)	absa	anggold	barworld	bidvest	gfields	har
3.00%	30.00%	0.71%	30.00%	30.00%	4.42%	4.87%
5.00%	30.00%	3.84%	30.00%	30.00%	4.13%	2.04%
7.50%	30.00%	6.93%	27.85%	30.00%	3.54%	1.68%
10.00%	30.00%	9.93%	27.22%	30.00%	2.52%	0.33%
12.50%	30.00%	12.27%	21.18%	30.00%	5.30%	1.26%
15.00%	30.00%	14.01%	18.52%	30.00%	6.75%	0.72%
17.50%	30.00%	16.64%	22.57%	30.00%	0.79%	0.00%
20.00%	30.00%	21.16%	16.65%	30.00%	2.19%	0.00%
22.50%	30.00%	23.74%	11.05%	30.00%	5.21%	0.00%
25.00%	30.00%	25.78%	7.18%	30.00%	7.05%	0.00%
30.00%	30.00%	28.44%	16.19%	25.37%	0.00%	0.00%
40.00%	29.32%	30.00%	0.00%	0.00%	22.33%	18.36%

Table 6.5: 95% Efficient Frontier

Risk (w)	% Maximized Return	% Actual Portfolio Return	Error
3.00%	0.20%	-2.26%	2.46%
5.00%	0.41%	-1.99%	2.40%
7.50%	0.76%	-1.51%	2.27%
10.00%	1.02%	-1.15%	2.17%
12.50%	1.24%	-0.92%	2.16%
15.00%	1.38%	-0.54%	1.91%
17.50%	1.53%	-0.23%	1.76%
20.00%	1.64%	0.23%	1.41%
22.50%	1.73%	0.36%	1.37%
25.00%	1.83%	0.58%	1.25%
30.00%	2.02%	1.23%	0.79%
40.00%	1.98%	2.81%	-0.83%

	99% CVaR Portfolio					
Risk (w)	absa	anggold	barworld	bidvest	gfields	har
3.00%	30.00%	0.00%	30.00%	30.00%	5.89%	4.11%
5.00%	30.00%	1.54%	30.00%	30.00%	4.91%	3.56%
7.50%	30.00%	4.37%	30.00%	30.00%	4.02%	1.62%
10.00%	30.00%	6.91%	30.00%	30.00%	3.09%	0.00%
12.50%	30.00%	9.35%	30.00%	30.00%	0.65%	0.00%
15.00%	27.86%	11.32%	30.00%	30.00%	0.83%	0.00%
17.50%	26.27%	13.73%	30.00%	30.00%	0.00%	0.00%
20.00%	30.00%	15.46%	23.47%	30.00%	1.07%	0.00%
22.50%	24.77%	17.63%	29.38%	27.96%	0.27%	0.00%
25.00%	20.99%	20.22%	28.79%	30.00%	0.00%	0.00%
30.00%	30.00%	24.55%	23.66%	21.80%	0.00%	0.00%
40.00%	30.00%	30.00%	10.56%	3.47%	16.11%	9.86%

Table 6.6: 99% Efficient Frontier

			% Allocation Markowitz Model					
Desired Return	Actual %Ret.	Error	absa	anggold	barworld	bidvest	gfields	har
0.25%	0.45%	-0.20%	23.19%	8.76%	23.68%	30.00%	14.37%	0.00%
0.30%	0.43%	-0.13%	23.51%	7.57%	23.12%	30.00%	15.80%	0.00%
0.35%	0.29%	0.06%	24.45%	3.59%	21.33%	30.00%	19.46%	1.17%
0.40%	0.19%	0.21%	25.38%	0.00%	19.00%	30.00%	22.81%	2.81%
0.45%	0.74%	-0.29%	27.33%	0.00%	19.66%	23.80%	26.30%	2.92%
0.50%	1.25%	-0.75%	29.29%	0.00%	20.38%	17.53%	29.77%	3.03%
0.55%	1.54%	-0.99%	30.00%	0.00%	22.64%	10.47%	30.00%	6.90%
0.60%	1.78%	-1.18%	30.00%	0.00%	25.48%	3.43%	30.00%	11.09%
0.65%	2.02%	-1.37%	30.00%	0.00%	23.52%	0.00%	30.00%	16.48%
0.70%	2.28%	-1.58%	30.00%	0.00%	16.97%	0.00%	30.00%	23.03%
0.75%	2.50%	-1.75%	30.00%	0.00%	10.43%	0.00%	30.00%	29.57%
0.80%	2.51%	-1.71%	30.00%	0.00%	10.00%	0.00%	30.00%	30.00%

Table 6.7: Markowitz Frontier

			% Allocation CAPM Model					
Desired Return	Actual %Ret.	Error	absa	anggold	barworld	bidvest	gfields	har
0.25%	0.30%	-0.05%	13.22%	8.69%	29.39%	30.00%	10.05%	8.65%
0.30%	0.30%	0.00%	13.22%	8.69%	29.39%	30.00%	10.05%	8.65%
0.35%	0.50%	-0.15%	12.23%	7.50%	26.80%	30.00%	12.83%	10.64%
0.40%	0.73%	-0.33%	11.20%	6.20%	24.15%	29.53%	16.00%	12.92%
0.45%	1.02%	-0.57%	11.29%	5.74%	23.92%	25.60%	18.59%	14.86%
0.50%	1.30%	-0.80%	11.37%	5.26%	23.71%	21.67%	21.19%	16.81%
0.55%	1.56%	-1.01%	11.44%	4.78%	23.50%	17.74%	23.78%	18.75%
0.60%	1.81%	-1.21%	11.52%	4.30%	23.30%	13.81%	26.37%	20.70%
0.65%	2.04%	-1.39%	11.60%	3.82%	23.09%	9.88%	28.97%	22.64%
0.70%	2.21%	-1.51%	12.18%	3.27%	23.73%	4.76%	30.00%	26.07%
0.75%	2.27%	-1.52%	19.87%	0.00%	20.13%	0.00%	30.00%	30.00%
0.80%	2.51%	-1.71%	30.00%	0.00%	10.00%	0.00%	30.00%	30.00%

Table 6.8: CAPM Frontier

Chapter 7

Conclusion

Based on the findings of the study, the following conclusions may be drawn:

We found that the marginal distributions for the individual stocks are best modelled using a t-distribution with t=7 and that the dependency between the stocks ~~are~~ best described by a T-Copula with v=3.

The scenario generation is the most computationally time consuming part of the CVaR model, the model stabilises for VaR and CVaR at around 1500 scenarios for the portfolio of chosen stocks and both confidence intervals. Generating 1500 scenarios for 6 stocks takes 5 seconds and is therefore not computationally intensive, when considering the time taken.

For the backtesting module of the study, we find that the CVaR value is always greater than the monetary risk measure calculated using VaR/alpha from the CVaR model, Non-parametric/Monte-Carlo VaR, and parametric VaR. The CVaR monetary risk value will therefore always be the indicator which is exceeded the least by an actual portfolio loss.

During the period we backtested (2000-04-22 to 2003-01-25) the number of times the risk measures are exceeded are as follows; for the 95% confidence level CVaR = 5%, VaR/alpha and Non-parametric/Monte-Carlo VaR = 14% and parametric VaR = 8%, the 99% confidence level results show CVaR = 1% while the other VaR indicators are all exceeded 4% of the time.

Scenarios	Seconds	Minutes
500	140.74	2.35
1000	170.50	2.84
2000	235.54	3.93
4000	348.37	5.81
6000	730.52	12.18
8000	1084.78	18.08
10000	1459.97	24.33

Table 7.1: Scenario Time Table

Considering the time taken to generate scenarios and the fact that we used 2000 scenarios to backtest our model for 100 weeks, we calculate that it took about $4 \times 100 = 400min = 6hrs40min$ to generate the scenarios. The CVaR optimisation algorithm takes about $15sec$ to optimise for each backtested period, giving us a total time of about $7hrs05min$ for our backtesting.

The non-parametric/Monte Carlo VaR does not need the CVaR optimiser and would therefore take $6hrs40min$ and the parametric VaR only needs the historical covariance matrices, which are not that time consuming to calculate, and took about $40min$ to backtest.

We can therefore conclude that CVaR is the most accurate but time consuming risk measure during our period of investigation.

The risk measure envelopes plotted for CVaR and alpha/VaR show an increase in size during the mid-point of the period of investigation. After plotting the EWMA variance of the stocks in the portfolio, we find that there is an increase in variance for four of the six stocks at the onset of the increase in envelope size; this indicates that our model does take into effect the changes in variance of the underlying instruments.

The CVaR efficient frontier for both confidence levels has a concave shape and levels off at $R_{CVaR} = 30\%$, this means that the investor must be willing

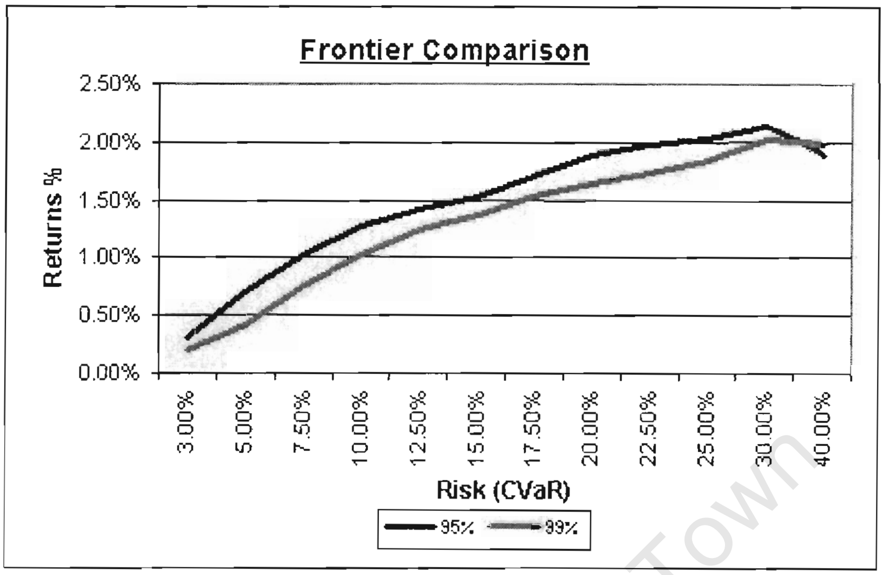

Figure 7.1: Frontier Comparison

to lose 30% of the portfolio in order to obtain the maximum possible one week return.

Figure 7.1 plots the two CVaR frontiers on one graph, from this graph we can conclude that the 99% frontier is a more conservative predictor of returns than the 95% frontier due to its lower maximised returns for the same levels of risk.

Comparing the tracking errors of the returns for the four efficient frontier portfolios we find that the CAPM Model has the lowest error followed by the Markowitz model, 99% CVaR model and the 95% CVaR model with the highest tracking error.

The actual portfolio returns for the four frontiers show that actual portfolio returns outperform the CAPM and Markowitz frontier portfolios more often than the CVaR models. The advantage of this is that the client would have received a return greater than what was predicted by the CAPM and Markowitz models, but the disadvantage is that the true portfolio returns do not really fall within the frontiers generated by these two models, while most of the actual portfolio returns are contained within the CVaR frontiers.

Chapter 8

Recommendations for Future Work

Based on the findings and conclusions of the study, the following recommendations for future work are made:

Effects of Volatility Models on CVaR Optimization

The effect of different volatility models on the CVaR model should be studied. This study will help determine how much of the portfolio variance is actually captured by the CVaR model, and which volatility model will ensure that the CVaR model best captures this volatility.

Scenario Stability For Varying Portfolio Size

A study can be conducted on how the stability of the model changes when we increase the number of instruments in the portfolio. This will determine how computationally intensive the model becomes as we increase the number of instruments in the portfolio.

Using Non-linear Instruments in CVaR Portfolio Optimization

The inclusion of non-linear instruments, such as options and interest rate derivatives available in the South African market, in the portfolio and their effects on the CVaR envelope when backtested should be studied. This study can be used to determine the feasibility of the CVaR model for hedged portfolios.

Chapter 9

Summary Chapter

The study has shown the feasibility of using a monetary risk measure to optimise a portfolio of South African stocks over a one-week period. The monetary risk measure tries to give an answer to the question of how much capital should be set aside to hedge down-side risk and we motivate the use of a coherent risk measure, namely conditional value at risk (CVaR). Coherent risk measures have the properties of relevance, monotonicity, positive homogeneity, sub additivity and translation invariance.

VaR is not a coherent risk measure because it does not have the property of sub additivity, we show this with an example, which means the VaR of a portfolio of two different securities can be more than the sum of the individual VaR of each security, and this lack of sub additivity defeats the aim of reducing risk by diversifying a portfolio.

The dependence structure of the instruments in the portfolio is modelled using a t-copula instead of a multivariate normal distribution due to our findings that this dependence structure better describes the historical dependence between the instruments.

After the introduction of CVaR and its properties we develop an algorithm for portfolio optimisation, this algorithm is programmed in visual basic for applications and executed in the Excel environment mainly because we are searching for a model that is easy to implement and modify in an environment that is easily available to most computer users.

Having backtested the model over 100 successive weeks for two confidence levels, 95% and 99%, we find that the CVaR risk measure is exceeded by the actual portfolio losses 5% and 1% of the time, for the respective confidence levels, during the 100 weeks. Comparing this to the VaR risk measures we find that the VaR risk measures are exceeded more than the CVaR risk measures for both confidence levels, the VaR risk measures are; parametric VaR, non-parametric VaR and VaR calculated from the CVaR model.

Using EWMA volatility method to calculate volatilities we find that the CVaR model shows the capability to capture increases in volatility of the underlying instruments in the portfolio, we see this with an increase in the size of the CVaR envelopes just after there is an increase in the volatility of the stocks in the portfolio.

We then generate efficient frontiers using modern portfolio theory and the CVaR model, again for the two confidence levels, the risk R_{CVaR} for the CVaR frontier is a percentage of the original equally weighted portfolio we are willing to lose. We find that the maximised returns predicted by the CVaR frontier contain more of the actual portfolio returns, using the frontier allocations, than the CAPM and Markowitz frontiers.

CVaR is therefore a good monetary risk measure for a portfolio of stocks, even though the scenario generation is time consuming. The next step is to investigate the effects of different volatility models and non-linear instruments on the performance of the model.

Chapter 10

APPENDIX

10.1 Proof of Theorem 1

For equation 3.6 we assumed that $\vartheta(\mathbf{x},\alpha)$ is continuous with respect to α, this means that.

$$\begin{aligned} \mathbf{P}[f(\mathbf{x},\mathbf{y})=\alpha] &= 0 \\ \int_{f(\mathbf{x},\mathbf{y})=\alpha} p(\mathbf{y})dy &= 0 \end{aligned} \tag{10.1}$$

Lemma. Let us fix $\mathbf{x}$, and let $H(\alpha)=\int_{y\in\mathbf{R}^m} h(\alpha,\mathbf{y})p(\mathbf{y})dy$ where $h(\alpha,\mathbf{y}) = [f(\mathbf{x},\mathbf{y})-\alpha]^+$. Then H is a convex continuously differentiable function with derivative.

$$H'\alpha = \vartheta(\mathbf{x},\alpha) - 1 \tag{10.2}$$

Proof.

$$\begin{aligned} H(\alpha) &= \int_{y\in\mathbf{R}^m} h(\alpha,\mathbf{y})p(\mathbf{y})dy \\ &= \int_{f(\mathbf{x},\mathbf{y})\geq\alpha} [f(\mathbf{x},\mathbf{y})-\alpha]p(\mathbf{y})dy \\ &= \int_{f(\mathbf{x},\mathbf{y})\geq\alpha} f(\mathbf{x},\mathbf{y})p(\mathbf{y})dy - \int_{f(\mathbf{x},\mathbf{y})\geq\alpha} \alpha p(\mathbf{y})dy \\ &= f(\mathbf{x},\mathbf{y})\int_{f(\mathbf{x},\mathbf{y})\geq\alpha} p(\mathbf{y})dy - \alpha[1-\int_{f(\mathbf{x},\mathbf{y})\leq\alpha} p(\mathbf{y})dy] \\ &\textit{then} \\ H'(\alpha) &= 0 - [1-\int_{f(\mathbf{x},\mathbf{y})\leq\alpha} p(\mathbf{y})dy] \\ &= \vartheta(\mathbf{x},\alpha)-1 \end{aligned}$$

Using **Lemma1**, we see that $F_\zeta(\mathbf{x},\alpha)$, equation (3.10) is convex and continuously differentiable with derivative.

$$\frac{\partial}{\partial\alpha}F_\zeta(\mathbf{x},\alpha) = 1 + (1-\zeta)^{-1}[\vartheta(\mathbf{x},\alpha) - 1] = (1-\zeta)^{-1}[\vartheta(\mathbf{x},\alpha) - \zeta]$$

The values of α for which $F_\zeta(\mathbf{x},\alpha)$ obtains its minimum are when $\vartheta(\mathbf{x},\alpha) - \zeta = 0$ these values of α are also the ones in the set $A_\zeta(\mathbf{x})$ and form a nonempty closed interval on $\mathbb{R}$ [1].We therefore have.

$$\min_{\alpha\in\mathbf{R}} F_\zeta(\mathbf{x},\alpha) = F_\zeta(\mathbf{x},\alpha_\zeta(\mathbf{x})) = \alpha_\zeta(\mathbf{x}) + (1-\zeta)^{-1}\int_{y\in\mathbf{R}^m}[f(\mathbf{x},\mathbf{y}) - \alpha_\zeta(\mathbf{x})]^+ p(\mathbf{y})dy$$

The integral on the right equals.

$$\int_{f(\mathbf{x},\mathbf{y})\geq\alpha_\zeta(\mathbf{x})} f(\mathbf{x},\mathbf{y})p(\mathbf{y})dy - \alpha_\zeta(\mathbf{x})\int_{f(\mathbf{x},\mathbf{y})\geq\alpha_\zeta(\mathbf{x})} p(\mathbf{y})dy$$

From the definition of equation (3.8), the first integral is equal to (1-ζ)$\Phi_\zeta(\mathbf{x})$ and from equation (3.6) the second integral is $1-\vartheta(\mathbf{x},\alpha_\zeta(\mathbf{x}))$, and since $\vartheta(\mathbf{x},\alpha_\zeta(\mathbf{x})) = \zeta$ we have

$$\min_{\alpha\in\mathbf{R}} F_\zeta(\mathbf{x},\alpha) = \alpha_\zeta(\mathbf{x}) + (1-\zeta)^{-1}[(1-\zeta)\Phi_\zeta(\mathbf{x}) - \alpha_\zeta(\mathbf{x})(1-\zeta)] = \Phi_\zeta(\mathbf{x})$$

This proves equation (3.11) and **Theorem1**

10.2 Equation 4.3 linear regression coefficients

Here we simply minimise $\mathbb{E}[(Y-(aX+b))^2]$

$$\begin{aligned}
& \mathbb{E}[(Y-(aX+b))^2] \\
= \; & \mathbb{E}[Y^2 + a^2X^2 + b^2 + 2abX - 2aXY - 2bY] \\
= \; & \mathbb{E}[Y^2] + a^2\mathbb{E}[X^2] + b^2 + 2ab\mathbb{E}[X] - 2a\mathbb{E}[XY] - 2b\mathbb{E}[Y]
\end{aligned}$$

to minimise the above equation

[1] $\vartheta(\mathbf{x},\alpha)$ is continuous and non decreasing on $\mathbb{R}$ with limits 1 and 0 for $\alpha = \infty$ and $\alpha = -\infty$ respectively

$$
\begin{aligned}
\frac{\partial}{\partial a} &= 2a\mathbb{E}[X^2] + 2b\mathbb{E}[X] - 2\mathbb{E}[XY] = 0 \\
a\mathbb{E}[X^2] &= \mathbb{E}[XY] - b\mathbb{E}[X]...(1) \\
\frac{\partial}{\partial b} &= 2b + 2a\mathbb{E}[X] - 2\mathbb{E}[Y] = 0 \\
b &= \mathbb{E}[Y] - a\mathbb{E}[X]...(2), \ \textit{subst into } (1) \\
a\mathbb{E}[X^2] &= \mathbb{E}[XY] - (\mathbb{E}[Y] - a\mathbb{E}[X])\mathbb{E}[X] \\
&= \mathbb{E}[XY] - \mathbb{E}[Y]\mathbb{E}[X] + a\mathbb{E}[X]\mathbb{E}[X] \\
a(\mathbb{E}[X^2] - \mathbb{E}[X]\mathbb{E}[X]) &= \mathbb{E}[XY] - \mathbb{E}[Y]\mathbb{E}[X] \\
a\sigma^2[X] &= Cov[X,Y] \\
a &= \frac{Cov[X,Y]}{\sigma^2[X]}
\end{aligned}
$$

So for a perfect linear regression of Y on X, $Y = aX + b$ we have constants:

$$
\begin{aligned}
a &= \frac{Cov[X,Y]}{\sigma^2[X]} \\
b &= \mathbb{E}[Y] - a\mathbb{E}[X]
\end{aligned}
$$

10.3 Correlation under linear Transformations

If we transform two real valued random variables X and Y under positive affine transformations $U = \alpha X + \beta$ and $Z = \gamma Y + \delta$ then $\rho(\alpha X + \beta, \gamma Y + \delta)$:

$$
= \frac{\mathbb{E}[(\alpha X+\beta)(\gamma Y+\delta)] - \mathbb{E}[(\alpha X+\beta)]\mathbb{E}[(\gamma Y+\delta)]}{\sqrt{(\mathbb{E}[(\alpha X+\beta)^2] - \mathbb{E}[(\alpha X+\beta)]\mathbb{E}[(\alpha X+\beta)])(\mathbb{E}[(\gamma Y+\delta)^2] - \mathbb{E}[(\gamma Y+\delta)]\mathbb{E}[(\gamma Y+\delta)])}} \tag{10.3}
$$

for the top part we have.

$$
\begin{aligned}
& \mathbb{E}[(\alpha X + \beta)(\gamma Y + \delta)] - \mathbb{E}[(\alpha X + \beta)]\mathbb{E}[(\gamma Y + \delta)] \\
= \ & \mathbb{E}[\alpha\gamma XY + \alpha\delta X + \beta\gamma Y + \beta\delta] - (\beta + \alpha\mathbb{E}[X])(\delta + \gamma[Y]) \\
= \ & \alpha\gamma[\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]] \\
= \ & \alpha\gamma Cov[X,Y]
\end{aligned}
$$

for the first part of the bottom part we have.

$$\begin{aligned}
& \mathbb{E}[(\alpha X+\beta)^2]-\mathbb{E}[(\alpha X+\beta)]\mathbb{E}[(\alpha X+\beta)] \\
= & \mathbb{E}[\alpha^2 X^2+\beta^2+2\alpha\beta X]-(\beta+\alpha\mathbb{E}[X])^2 \\
= & \alpha^2(\mathbb{E}[X^2]-(\mathbb{E}[X])^2) \\
= & \alpha^2\sigma^2[X] \quad similarly \\
& \mathbb{E}[(\gamma Y+\delta)^2]-\mathbb{E}[(\gamma Y+\delta)]\mathbb{E}[(\gamma Y+\delta)] \\
= & \gamma^2\sigma^2[Y]
\end{aligned}$$

So equation (10.3) equals:

$$\begin{aligned}
& \frac{\alpha\gamma Cov[X,Y]}{\sqrt{\alpha^2\sigma^2[X]\gamma^2\sigma^2[Y]}} \\
= & sgn(\alpha\cdot\gamma)\rho(X,Y)
\end{aligned}$$

10.4 Theorem 5.1

Under smooth conditions of $f(x/\theta)$ the MLE $\hat{\theta}$ is consistent[2]

Proof.

Let θ_0 be the true value of θ for $f(x/\theta)$, we have to show that $\hat{\theta} \xrightarrow{P} \theta_0$ as $n \to \infty$. For a sample of i.i.d. random variables $(X_1, ..., X_n)$ with density functions $f(x/\theta)$, the MLE maximises

$$\begin{aligned}
\ell(\theta) &= logL(\theta) \\
&= \sum_{i=1}^{n} log[f(X_i/\theta)]
\end{aligned}$$

Therefore it can also be used to maximise

$$\frac{1}{n}\ell(\theta) = \frac{1}{n}\sum_{i=1}^{n} log[f(X_i/\theta)]$$

[2] An estimate of θ based on a sample sise of n is said to be consistent if the estimate $\hat{\sigma}_n$ converges in probability to θ as we increase n.

$$P(|\hat{\theta}_n - \theta| > \epsilon) \to 0 \quad as \quad n \to \infty \tag{10.4}$$

for a small finite ϵ

Since the X_i are i.i.d. so are the $log[f(X_i/\theta)]$, and by the Weak Law of Large Numbers (WLLN) the right hand side of the equality converges to the mean (expected value) as $n \to \infty$ and we have.

$$\frac{1}{n}\ell(\theta) \xrightarrow{P} E[logf(X/\theta)] \quad as \quad n \to \infty$$

Therefore the $\hat{\theta}$ that maximises $\frac{\ell(\theta)}{n}$ should be a close approximation to the θ that maximises $E[logf(X_i/\theta)]$. So let us maximise $E[logf(X/\theta)]$ under the smoothness assumptions of $f(x/\theta)$.

$$\begin{aligned}
\frac{\partial E[logf(X/\theta)]}{\partial\theta} &= \frac{\partial}{\partial\theta}\int logf(x/\theta)f(x/\theta_0)dx \\
&= \int \frac{\partial}{\partial\theta}logf(x/\theta)f(x/\theta_0)dx \\
&= \int \frac{1}{f(x/\theta)}\frac{\partial f(x/\theta)}{\partial\theta}f(x/\theta_0)dx
\end{aligned}$$

We can change the order of integration because $f(x/\theta)$ is smooth. Now let us set $\theta = \theta_0$ we get

$$\begin{aligned}
\left[\frac{\partial E[logf(X/\theta)]}{\partial\theta}\right]_{\theta=\theta_0} &= \int \frac{1}{f(x/\theta_0)}\frac{\partial f(x/\theta_0)}{\partial\theta}f(x/\theta_0)dx \\
&= \int \frac{\partial f(x/\theta_0)}{\partial\theta}dx \\
&= \frac{\partial}{\partial\theta}\int f(x/\theta_0)dx \\
&= \frac{\partial}{\partial\theta}(1) = 0
\end{aligned}$$

and we therefore have

$$\hat{\theta} \xrightarrow{P} \theta_0 \quad as \quad n \to \infty$$

10.5 Theorem 5.2.

If $f(x/\theta)$ is smooth then the MLE $\hat{\theta}$ converges to the normal distribution as $n \to \infty$.

$$\hat{\theta} \to N(\theta_0, \frac{1}{nI(\theta_0)}) \quad as \quad n \to \infty$$

where θ_0 is the true value of parameter θ and $I(\theta)$ the asymptotic variance.

Before we prove this theorem we will need the following lemma.

Lemma. If

$$I(\theta) = E\left[\frac{\partial}{\partial\theta} log f(X/\theta)\right]^2 \tag{10.5}$$

Bartlett identity

then under the smoothness assumptions of $f(x/\theta)$.

$$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2} log f(X/\theta)\right] \tag{10.6}$$

Quick proof, we know that $\int f(x/\theta)dx = 1$ so

$$\frac{\partial}{\partial\theta}\int f(x/\theta)dx = 0$$

and since

$$\begin{aligned}
\frac{\partial}{\partial\theta} f(x/\theta) &= \left[\frac{\partial}{\partial\theta} log f(x/\theta)\right] f(x/\theta) \quad we \quad have \\
0 &= \frac{\partial}{\partial\theta}\int f(x/\theta)dx = \int \frac{\partial}{\partial\theta} f(x/\theta)dx \\
&= \int \left[\frac{\partial}{\partial\theta} log f(x/\theta)\right] f(x/\theta)dx
\end{aligned}$$

Taking second derivatives

$$\begin{aligned}
0 &= \frac{\partial}{\partial\theta}\int \left[\frac{\partial}{\partial\theta} log f(x/\theta)\right] f(x/\theta)dx \\
&= \int \left[\frac{\partial^2}{\partial\theta^2} log f(x/\theta)\right] f(x/\theta)dx + \int \left[\frac{\partial}{\partial\theta} log f(x/\theta)\right] \frac{\partial}{\partial\theta} f(x/\theta)dx \\
&= \int \left[\frac{\partial^2}{\partial\theta^2} log f(x/\theta)\right] f(x/\theta)dx + \int \left[\frac{\partial}{\partial\theta} log f(x/\theta)\right]^2 f(x/\theta)dx \\
&= E\left[\frac{\partial^2}{\partial\theta^2} log f(x/\theta)\right] + E\left[\frac{\partial}{\partial\theta} log f(x/\theta)\right]^2
\end{aligned}$$

And the result follows. Note from the proof of **Theorem 5.1.**, if we have $C = \frac{\partial}{\partial\theta} logf(x/\theta)$ then $E(C) = 0$ and we have.

$$VaR(C) = E(C^2) = E\left[\frac{\partial}{\partial\theta} logf(X/\theta)\right]^2 = I(\theta)$$

Proof of Theorem 5.2

The Taylor series expansion around θ_0 we get

$$\begin{aligned} \ell'(\hat{\theta}) &\approx [\ell'(\theta)]_{\theta=\theta_0} + (\hat{\theta} - \theta_0)[\ell''(\theta)]_{\theta=\theta_0} + \ldots \\ &\approx \ell'(\theta_0) + (\hat{\theta} - \theta_0)\ell''(\theta_0) + \ldots \end{aligned}$$

But $\ell'(\hat{\theta}) = 0$ since $\hat{\theta}$ is the MLE. so

$$(\hat{\theta} - \theta_0) \approx \frac{-\ell'(\theta_0)}{\ell''(\theta_0)} \quad or \tag{10.7}$$

$$\sqrt{n}(\hat{\theta} - \theta_0) \approx \frac{-n^{-1/2}\ell'(\theta_0)}{n^{-1}\ell''(\theta_0)} \tag{10.8}$$

Let us consider the denominator.

$$n^{-1}\ell''(\theta_0) = \frac{1}{n}\sum_{i=1}^{n}[\frac{\partial^2}{\partial\theta^2} logf(X_i/\theta_0)]$$

The denominator is the sum of i.i.d random variables and by the WLLN converges in probability to it's mean.

$$E\left[\frac{\partial^2}{\partial\theta^2} logf(X_i/\theta_0)\right] = -I(\theta_0)$$

The denominator is therefore a constant. The expectation and variance of the numerator are.

$$\begin{aligned}
E[n^{-1/2}\ell'(\theta_0)] &= n^{-1/2}\sum_{i=1}^{n} E[logf(x_i/\theta_0)] \\
&= n^{-1/2}\sum_{i=1}^{n} 0 \\
&= 0 \\
VaR[n^{-1/2}\ell'(\theta_0)] &= \frac{1}{n}\sum_{i=1}^{n} E\left[\frac{\partial}{\partial\theta}logf(X_i/\theta_0)\right]^2 \\
&= \frac{1}{n} nI(\theta_0) \\
&= I(\theta_0) \quad a \quad constant
\end{aligned}$$

We therefore have

$$\begin{aligned}
\sqrt{n}(\hat{\theta}-\theta_0) &\approx \frac{-n^{-1/2}\ell'(\theta_0)}{-I(\theta_0)} \\
&= \frac{n^{-1/2}\ell'(\theta_0)}{I(\theta_0)} \quad for \quad large \quad n
\end{aligned}$$

and

$$\begin{aligned}
E[\sqrt{n}(\hat{\theta}-\theta_0)] &= \frac{1}{I(\theta_0)} E[n^{-1/2}\ell'(\theta_0)] \\
&= 0 \\
VaR[\sqrt{n}(\hat{\theta}-\theta_0)] &= \frac{1}{[I(\theta_0)]^2} VaR[n^{-1/2}\ell'(\theta_0)] \\
&= \frac{1}{[I(\theta_0)]^2} I(\theta_0) \\
&= \frac{1}{I(\theta_0)}
\end{aligned}$$

So Since

$$\sqrt{n}(\hat{\theta}-\theta_0) \approx \frac{n^{-1/2}\ell'(\theta_0)}{I(\theta_0)} \quad for \quad large \quad n$$

and

$$n^{-1/2}\ell'(\theta_0) = n^{-1/2}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}logf(X_i/\theta_0)$$

$n^{-1/2}\ell'(\theta_0)$ is therefore the sum of n i.i.d. random variables with the same mean (0) and variance $(I(\theta_0))$ at $\theta = \theta_0$, and from the Central Limit Theorem it follows that it converges to the normal distribution as $N \to \infty$, so.

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta_0) \quad &\xrightarrow{d} \quad N(0, \frac{1}{I(\theta_0)}) \quad as \quad n \to \infty \\ or \quad \hat{\theta} \quad &\xrightarrow{d} \quad N(\theta_0, \frac{1}{nI\theta_0}) \quad as \quad n \to \infty \end{aligned}$$

And the theorem is proved

Bibliography

[1] P.Artzner, F.Delbaen, J.Eber, D.Heath 1998:*Coherent Measures of Risk.* (http://gloriamundi.org/picsresources/adeh.pdf)

[2] C.Acerbi, R.Tasche 2001:*Expected Shortfall: a natural coherent alternative to Value at Risk.* (http://gloriamundi.org/picsresources/expshortfall.pdf)

[3] C.Acerbi, R.Tasche 2002:*On the coherence of Expected Shortfall.* (http://gloriamundi.org/picsresources/cadt.pdf)

[4] A.Gaivoronski, G.Pflug :*Value at Risk in Portfolio Optimization: Properties and Computational Approach.* (http://gloriamundi.org/picsresources/ggp1.pdf)

[5] T.Rockafellar, S.Uryasev 2001":*Conditional Value-At-Risk For General Loss Distributions.* (http://gloriamundi.org/picsresources/rtrsu.pdf)

[6] M.Ammann and C.Reich 2001:*Value-at-Risk for Nonlinear Financial Instrument-Linear Approximation or Full Monte-Carlo?.* (http://www.gloriamundi.org/picsresources/macr.pdf)

[7] B. Vas de Melo Mendez, R.M.de Souza.(2004):*Measuring financial risks with Copulas.* International Review of Financial Analysis, Volume 13, Issue 1, Spring 2004, Pages 27-45

[8] S. Benati.(2003): *The optimal portfolio problem with coherent risk measure constraints.* European Journal of Operational Research, Volume 150, Issue 3, 1 November 2003, Pages 572-584

[9] C.G.Troskie 2003:*Multivariate analysis and Regression Time Series.* University of Cape Town Lecture notes

[10] C.G.Troskie. 2003: *Statistical Inference.* University of Cape town Lecture Notes

[11] C.G.Troskie. 2003: *Multivariate Statistical Modeling.* University of Cape town Lecture Notes

[12] Z.Bode, A.Kane and A.J.Marcus.(2002): *Investments, International Edition.* McGraw-Hill, New York

[13] Eric Bouyé. (2000): *Copulas for Finance A Reading Guide and Some Applications.*(http://gro.creditlyonnais.fr/content/wp/copula-survey.pdf)

[14] S.Demarta, A.McNeil 2004:*The t Copula and Related Copulas.* (http://www.math.ethz.ch/~mcneil/ftp/tCopula.pdf)

[15] Patrick Cheridito, Freddy Delbaen and Michael Kupper.(2003): *Coherent and Convex monetary risk measures for bounded càdlàg processes.* (http://www.gloriamundi.org/picsresources/cdk.pdf)

[16] A. Di Clemente and C.Romano. (2003): *Measuring Portfolio Value-At-Risk by a Copula-EVT based approach.* (http://www.gloriamundi.org/picsresources/adcr4.pdf)

[17] Paul Embrechts, Alexander McNeil and Daniel Straumann.(1999): *Correlation and dependence in risk management: Properties and Pitfalls.* (http://www.ccfz.ch/files/Embrechtspitfalls.pdf)

[18] R.Dobbins, S.Witt and J.Fielding .(1996):*Portfolio Theory and Investment Management* , 2^{nd} *Edition.*Blackwell.Oxford.

[19] J.C.Hull(2000):*Options, Futures and Other Derivatives,* 4^{th} *Edition.* Prentice Hall International limited, London.

[20] P. Jorion : *How Informative are Value-At-Risk Disclosures.* (http://www.edhec-risk.com/site_edhecrisk/public/research_news/choice/RISKReview104618841305081372)

[21] N.Larsen, H.Mausser, S.Uryasev. (2001): *Algorithms for Optimization of Value-At-Risk.*(http://www.ise.ufl.edu/uryasev/pubs.html)

[22] P.Embrechts, F.Lindskog and A.Mcneil. (2001): *Modelling Dependence with Copulas and Applications to Risk Management.* (http://www.risklab.ch/ftp/papers/DependenceWithCopulas.pdf)

[23] R.Tyrrell Rockafellar and Stanislav Uryasev.(1999):*Optimization of Conditional Value-at-Risk.* (http://www.ise.ufl.edu/uryasev/cvar.pdf)

[24] Claudio Romano: *Calibrating and Simulating Copula Functions: An Application to the Italian Stock Market.*(w3.uniroma1.it/cidem/Files/WPRomanodicembre02.pdf)

[25] Georg Ch. Pflug 2000:*Some remarks on the Value-At-Risk and the Conditional Value-At-Risk.* (http://www.gloriamundi.org/picsresources/gpfl.pdf)

[26] J.Rank 2000:*Copulas in Financial Risk Management.* (http://www.gloriamundi.org/picsresources/jr.pdf)

[27] G.J. Alexander, A.M. Baptista 2003:*CVaR as a Measure of Risk: Implications for Portfolio Selection.* (http://gloriamundi.org/picsresources/gjaamb.pdf)

[28] Riskmetrics-Technical Document, Fourth Edition 1996. (www.Riskmetrics.com)